Università degli Studi di Napoli Federico II

Ph.D. Program in
Information and Communication Technology for Health
XXXVII Cycle

Thesis for the Degree of Doctor of Philosophy

The Unscripted Encounter: Social Cues for Spontaneous Human-Robot Interactions

by

Francesco Vigni

Advisor: Prof. Silvia Rossi

Scuola Politecnica e delle Scienze di Base
Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione

Es el mejor de los buenos quien sabe que en esta vida
todo es cuestiön de medida: un poco más, algo menos.
Antonio Machado

To my father Alfiero Vigni
my mother Luz Margarita Frias Frias
my brother Alessandro Vigni

The Unscripted Encounter: Social Cues for Spontaneous Human-Robot Interactions

Ph.D. Thesis presented
for the fulfilment of the Degree of Doctor of Philosophy
in Information and Communication Technology for Health
by

Francesco Vigni

October 2024

Approved as to style and content by

Prof. Silvia Rossi, Advisor

Università degli Studi di Napoli Federico II
Ph.D. Program in Information and Communication Technology for Health
XXXVII cycle - Chairman: Prof. Daniele Riccio

http://icth.dieti.unina.it

Candidate's declaration

I hereby declare that this thesis submitted to obtain the academic degree of Philosophiæ Doctor (Ph.D.) in Information and Communication Technology for Health is my own unaided work, that I have not used other than the sources indicated, and that all direct and indirect sources are acknowledged as references.
Parts of this dissertation have been published in international journals and/or conference articles (see list of the author's publications at the end of the thesis).

Napoli, October 31, 2024

Francesco Vigni

Abstract

This thesis investigates the significance of non-verbal behaviours in promoting spontaneous Human-Robot Interaction (HRI). This research recognises the inherent challenge associated with modelling spontaneity and introduces a model known as the Spontaneous Interaction State Machine (SISM). This model underscores the significance of interaction state and context in the comprehension of social environments. Empirical studies involving diverse robotic platforms show that robots can effectively employ a variety of social cues, ranging from basic signals such as lights to intricate emotional expressions, to initiate and sustain interactions within dynamic social environments.

This thesis emphasizes two key aspects: the role of gaze and the importance of proximity in enhancing spontaneous HRI. The findings highlight the importance of gaze as a fundamental social cue, demonstrating that users' perception of a robot's social presence is significantly affected by its gazing behaviours. Additionally, proximity emerges as a crucial factor, with adaptive use of distance helping robots to respect personal space. This research underscores the adaptive capabilities of robots in modifying their behaviours in response to human emotional states, thereby enriching the interaction experience. A lightweight and modular engagement metric based on non-verbal behaviours such as gaze and proximity is presented and validated. The methodological contributions include tools aimed at improving the reliability of datasets and advancing the standardisation of research methodologies via software containerization.

As we move towards a world increasingly populated by social robots, the insights gained here promise to enhance the quality of HRI, fostering cooperation and adaptability to diverse real-world contexts. This research enhances our comprehension of spontaneous HRI and acts as a catalyst for breakthroughs that will influence the future of robotics in everyday life.

Keywords: Spontaneous HRI, Non-Verbal Behaviours, Robots' Social Cues, Measuring Engagement

Sintesi in lingua italiana

Questa tesi indaga l'importanza dei comportamenti non verbali nella promozione delle interazioni umano-robot spontanee. Questa ricerca riconosce la sfida intrinseca associata alla modellazione della spontaneità e introduce un modello noto come Spontaneous Interaction State Machine (SISM). Questo modello sottolinea l'importanza dello stato di interazione e del contesto nella comprensione dei comportamenti sociali. Studi empirici condotti su diverse piattaforme robotiche mostrano che i robot possono utilizzare efficacemente una varietà di segnali sociali, che vanno da segnali di base come le luci a complesse espressioni emotive, per avviare e sostenere interazioni in ambienti sociali dinamici.

Questa tesi pone l'accento su due aspetti chiave: il ruolo dello sguardo e l'importanza della prossemica nel migliorare le interazionie spontanee. I risultati evidenziano l'importanza dello sguardo come indizio sociale fondamentale, dimostrando che la percezione della presenza sociale di un robot da parte degli utenti è significativamente influenzata dal suo sguardo. Inoltre, la prossemica emerge come un fattore cruciale, con un uso adattivo della distanza che aiuta i robot a rispettare lo spazio personale. Questa ricerca sottolinea le capacità adattive dei robot nel modificare i loro comportamenti in risposta agli stati emotivi umani, arricchendo così l'esperienza di interazione. Viene presentata e validata una metrica di coinvolgimento leggera e modulare basata su comportamenti non verbali come lo sguardo e la prossemica. I contributi metodologici includono strumenti volti a migliorare l'affidabilità delle basi di dati e a far progredire la standardizzazione delle metodologie di ricerca attraverso la containerizzazione del software.

Mentre ci avviciniamo a un mondo sempre più popolato da robot sociali, le intuizioni acquisite promettono di migliorare la qualità delle interazioni umano-robot, favorendo la cooperazione e l'adattabilità a diversi contesti del mondo reale. Questa ricerca migliora la nostra comprensione delle interazioni spontanee con i robot e funge da catalizzatore per le scoperte che influenzeranno il futuro della robotica nella vita quotidiana.

Parole chiave: interazioni spontanee, comportamenti non verbali, segnali sociali dei robot, misurazione del coinvolgimento.

Funding

The research presented in this dissertation has been supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 955778.

PERSEO

European Training Network on
PErsonalized Robotics as SErvice Oriented
applications

MARIE CURIE ACTIONS

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Question everything. Learn something. Answer nothing.

Euripides

The report of Price Waterhouse Coopers from 2018 [77] highlights three main overlapping themes (or waves) about automation that will shape the job market by 2030. They refer to the three waves as the algorithm wave, augmentation wave, and autonomy wave. The augmentation wave, which focuses on automating repetitive tasks and exchanging information through dynamic technological support, is the most appropriate one to analyze when considering robots designed for social interactions. Social robots can fit this description, yet they are built to tackle challenges that go beyond performing a task, such as detecting and manipulating objects in a well-known environment. The keyword *social* stresses the social capabilities that these robotics artefacts are designed with. They can adapt to the ever-changing and intricate human social environment.

Imagine the busy environment of an international airport, where people are arriving from all around the world. Imagine a robot whose job it is to navigate through this busy crowd while providing travel information. Without much effort, imagine the unique communication challenges that this robot can encounter.

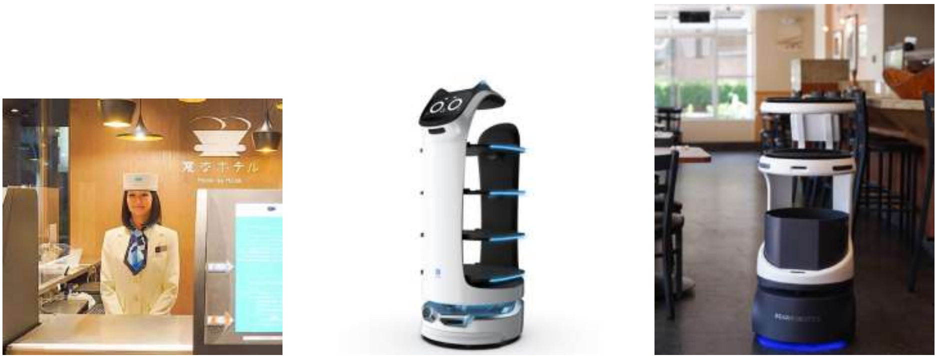
Take language, for example. If everyone speaks English, it might work most of the time, but a number of passengers will feel left out. Language or auditory barriers may be the cause. Rather than relying solely on speech, this robot could utilize non-verbal cues to proactively approach someone. It could fix its gaze on a particular traveler, moving gently toward them to signal its intent to initiate an interaction. This silent yet intentional approach can convey its Willingness to Engage (WtE), creating a moment of connection before determining the passenger's preferred language for communication. This scene can serve as an illustrative example of how robots can leverage non-verbal behaviours to initiate interactions, closely recalling similar interaction settings available in the literature [171, 193].

The goal of achieving long-term deployments of social robots, with the capability to autonomously conduct social interaction, is far from being reached when looking at the past and present tentatives. For instance, the Henn-na Hotel in Japan started operations in mid-2015 and is an outstanding example of the implementation of service robots in the front lines within the hospitality sector (see Figure 1.1a). It is known as the first fully automated hotel staffed entirely by robots, wherein guests do not engage with human employees at any point during their service experience.

Other examples of robots in the hospitality sector are given by the *BellaBot Pro* from Pudu Robotics [157] (see Figure 1.1b) and its more recent competitor the *Servi robot* from Bear Robotics [156] (see Figure 1.1c). These robots support food servings in restaurants, carrying meals to and from diners' tables. These are designed as tower of trays fixed on top of an Autonomous Mobile Robot (AMR) without robot arms. This lack, necessary for the pick-and-place task, results in the robots being responsible for merely transporting meals from one location to another in an indoor environment, i.e., a restaurant. A human is employed to complete servings by picking the dishes from and to the robot trays.

These examples of social robots' deployment for long-term interactions in social spaces offer great insights for understanding how these are perceived and to what extent are accepted by users. Despite the initial hype and use cases these robots were addressing, they also revealed limitations that can significantly impact real integration in social environments.

Considering the case of the android receptionist in the Henn-na hotels, after around 5 years of operation, the management decided to eliminate roughly fifty percent of their robotic workforce due to several negative experiences reported [59]. The inability of these robots to go beyond language



(a) Android robot for assisting check-ins in Henn-na Hotels.

(b) BellaBot Pro from Pudu Robotics.

(c) Servi robot from Bear Robotics.

Figure 1.1. Examples of robots employed in single contexts of the hospitality sector.

barriers, as well as the lack of effectively providing the assistance that was requested, caused frustration in many guests. After this attempt to integrate robots in the front desk, the hotel management decided to opt for a hybrid approach in which human personnel provide the first assistance and a robotic agent completes the procedure. A similar cooperation strategy can also be seen in the applications of *BellaBot Pro* or *Servi robot*. In this way, the hotel requires roughly 23% of the human personnel of a structure with similar requirements and has integrated the robots as collaborators in the check-in process rather than fully in charge of it.

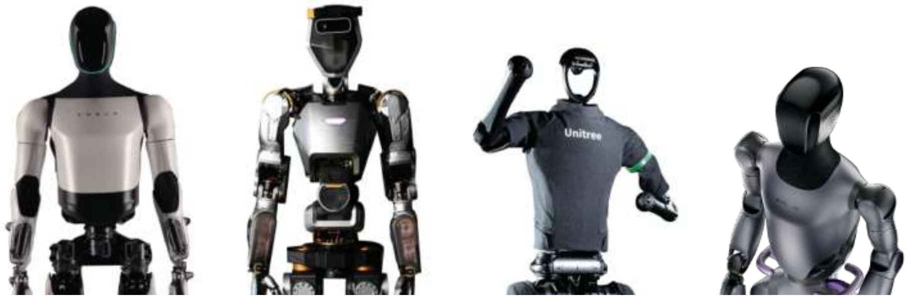
Regarding robots as servers in restaurants, Chen *et al.*[39] investigate the satisfaction of a restaurant's staff and diners with a survey. Their results suggest that, despite the robot taking longer for serving food (2 to 3 minutes compared to about one minute for a trained staff), diners were significantly more satisfied with the robot with respect to the human staff. Despite its daily tasks only involve bringing dishes in a restaurant to the tables, we can assimilate the role of the *Servi robot* as similar to serving staff, and it is likely that diners would spontaneously attempt to request items or services from the robot as well. This is a skill that experienced human staff master well; however, robots are still not capable of doing that.

In this sense, a spontaneous direct customer request to the robot might be ignored given its task-based design to swiftly move around the environment carrying items on its trays. Ignoring spontaneous requests from users, such as customers of a restaurant, could likely cause frustration and disappointment, leading to complaints towards the robot platform, and ultimately impacting its social acceptance.

- How should robots operate within social environments?
- How can robots conduct spontaneous interactions?

These questions conceal the rationale of instrumenting robots with human-like social skills. In other words, this is what is frequently referred to as the "humanisation of robots." Giger and colleagues in [68] refer to it as the effort to make robots to closely mimic human appearance and behaviours, including the display of humanlike cognitive and emotional states. They also highlight a possible way for improving Human-Robot Interaction (HRI) by using bottom-up approaches to build social robots. These consist of instrumenting robots with a combination of human features (e.g., limbs, lips, and eyebrows) and human-like behaviours (e.g., facial expressions, body gestures, tone of voice) robots regardless of their appearance.

One of the most significant advantages of humanising robots is the development of a mutual psychological understanding between the two parties during HRIs [176]. Drawbacks when humanizing robots are present. For example, Strait *et al.*[188] found that participants had more negative reactions to human-like robots compared to less human-like ones or even human agents. The study revealed that not only did participants rate these robots more negatively, but they also displayed greater avoidance of such encounters. Similarly, Waytz *et al.*[204] showed that people experienced stronger feelings of threat when robots were perceived as capable of replacing humans in emotion-oriented tasks, which are traditionally seen as human domains. This sense of threat was particularly pronounced when participants were informed that robots could outperform humans in both physical and mental tasks. However, this effect was less significant in cognitive-oriented tasks, which are perceived as more appropriate for robots.



Optimus Gen2 from Tesla.

Phoenix from Sanctuary AI.

H1 from Unitree.

Gr-2 from Fourier.

Figure 1.2. Recent examples of humanoid robots.

The questions introduced above are relevant regarless of the robot design. Recent years have seen increasing attention in robots with anthropomorphic features like the one in Figure 1.1a. Leading to the development of humanoid robots. Figure 1.2 shows a few recent examples of humanoid robots developed by the efforts of four different organisations worldwide. They are all designed with parts that closely resemble humans, such as a face, two arms, and two legs. Supporting this trend is also a report from the consulting company Root Analysis evaluating the market value of humanoids to reach to 243.4 billion USD by 2035 [7]. Another report shrinks the global market size for humanoid robots to the value of 66.0 billion USD by 2032 [86]. Finally, according to a report from Goldman Sachs Research, the market for these is expected to range between 4.2 and 37.8 billion USD by 2035 [153]. These values consider various levels of robot sophistication, from basic functionality to state of the art. From these, it can be grasped that robots with various levels of resemblance to humans are likely to have an impact in our daily lives. Robots, regardless their anthropomorphism, will impact our daily lives and will be asked to swiftly adapt to various interaction contexts.

The interaction context plays a role on how the robot shall behave. On this topic, the work of Menenez *et al.*[129] focuses on the role of context information in the interaction process between social robots and people by introducing a framework for context-based HRI. The rationale is that since we naturally adapt to different contexts, if we want robots among us, they shall do that as well. The authors define context as “the set of information that is relevant, affects or constrains how some action is taken without being at the centre of interest of the search or action.” The approach in Menenez *et al.*[129] has inspired this thesis in two separate ways.

First, the concept of *Context models* as a-priori data models the robots can exploit for the interaction. For instance, a robot is standing in front of a table and it is able to see a cup and a bottle of water. In this case, the *Context model* could be represented by the link between the cup and the bottle of water. Suggesting a possible action to the robot.

Second, it introduces for a robot the need of a “periodic process that operates in the background of the system” that can assess the context. This, is referred by Menenez and colleagues as a *context recognition system*. For example, a robot could measure the sound noise in a room and autonomously decide to increase the volume of its speaker to interact verbally with a person. This periodic process can be seen as a continuous awareness of the robot of its surroundings. Different surrounding, or social environment, require different operationalisation of *context recognition systems*. For instance, a robot teacher that has to maintain the attention of its student [60, 49, 88] shall use a different metric than a robot acting as a bartender that looks for which customer to serve [160]. The underlying assumption is that the engagement metric is a form of a *context recognition system*. A similar rationale is already tackled by Salam & Chetouani [167] in which authors hypothesised that “the definition of engagement varies in function of the context of the interaction”.

1.1 The Focus

The focus of this thesis is to investigate how social robots can capture dynamic nuances of our social environments and react with the available interaction interfaces in a spontaneous-like way. Particular attention is given to the non-verbal communication channel that can inform robots about surrounding humans, and it can also be used to communicate to them.

1.2 The Motivation

Two primary motivations drive this research. On one hand, the societal challenges that can be addressed with social robots are growing. Hospitality is just one of the possible impacted sectors. On the other hand, there is a novel *moon race* for developing robots with human-like aspects that can effectively address them. This highlights the need for robots to handle scenarios in which interactions can start in unscripted ways. Spontaneously. Their ability to "read the room" and understand subtle cues offered by surrounding humans as well as the context is likely to draw the line between the robots that will be discontinued after a trial period and the ones that will make a dent on how we inhabit our social environments.

Social robots will inevitably operate among us with designs and behaviours that will insinuate their capability of manage spontaneous interactions swiftly. For instance, Schulte *et al.*[175] considered spontaneous interaction those with users that were not previously trained to interact with the robot providing tour services in a museum. Ben-Youssef *et al.*[23] model spontaneous interactions by empowering participants the freedom to enter or leave the interaction. With a similar rationale, the work of Arreghini *et al.*[14] models as spontaneous interactions those exhibited by users physically approaching the robot. The authors developed a robotic service task of offering chocolate treats using an Long-Short Term Memory (LSTM) to capture the dynamic of passer-by people and offer chocolate to the one that is more likely to engage with the robot. When attempting to infer the robot's visual perspective, Zhao & Malle [215] model spontaneity as "an action taken by users without an explicit prompt."

Yet, this *spontaneous* dimension of interactions is still broadly defined. The main gap identified in the literature is that despite spontaneous interactions being expected to be prevalent within social robots, no unified way of modelling these, exists yet. For this reason, this thesis proposes a finite-state approach to identify possible interaction phases (or states) with a narrow focus on the significant social aspects of these.

1.3 Research Goal

The research goal of this thesis is to improve social perception techniques of robots, while focusing on spontaneous interactions. These are expected to happen frequently with robots around us and we can take advantage of non-verbal communication to establish how interactions unfold. This choice is made considering the heterogeneous population that social robots in public spaces will interact worldwide.

1.3.1 Contributions

The scientific contributions of this thesis emphasise the importance of abstracting the representation of a spontaneous HRI, aiming to build robots that can swiftly integrate into our social environments. This begins with equipping the robot with essential strategies to effectively display social cues. Following this, we explore how these cues initiate interactions and conclude by analyzing their role in sustaining interactions. Achieving these goals hinges on the robot's ability to accurately *measure* interaction, enabling it to adapt meaningfully within dynamic social contexts. For this reason, the last part of the thesis presents and validate a metric that exploits non-verbal cues to assess engagement.

The user studies reported in this thesis provide valuable insights with respect to physical and virtual interaction with robots. Overall, the user studies involved 438 healthy adults participants, of which most of them (N=268) interacted with a robot in the real world while a fewer portion conducted the study remotely (N=170). Particular attention is posed on the ecological validity of the studies. This is done by prioritizing user studies *in-the-wild* for which the conclusions are meant to be generalisable to the real world. That means testing the experimental conditions outside the constraints of the lab and directly in a real-life setting.

This thesis discusses the results from seven user studies (of which two conducted *in-the-wild*), one methodological improvement for assessing the reliability of datasets, and one explainable and lightweight engagement metric.

Technical contributions are also available as a result of the work presented here. These serves as the support to conduct the user studies as well as improving the reliability of datasets. With the aim of improving the reproducibility of software implementation in HRI, the implementations have adopted a Test-Driven Development (TDD) approach in which test cases are written *before* the actual implementation. As a result, most of the functionalities are safeguarded by programmatic tests.

1.4 Research Questions

This introduction chapter highlights the need for building social robots that are capable of conducting autonomously spontaneous interactions. It is important to consider that the design and the behaviour of a robot deployed in a social environment can be attributed social valence [174]. Running into the risk of developing a robot without controlling how it displays social cues.

The following research questions are formulated to systematically explore our conceptualisation of how spontaneous interactions are anticipated to occur and the robot's ability to assess interactions.

RQ1 How can robots display social cues?

In the context of social environments, robots are capable of displaying social cues that range from simple, familiar behaviours to more complex, emotion-driven expressions. The key focus here is how robots can effectively exhibit these cues. This leads to dissect **RQ1** into these subquestions:

RQ1.1 How can robots effectively display familiar, simple social cues?

RQ1.2 How can robots display complex social cues, such as emotions?

These questions aim to investigate the nuances of social signals a robot can display, starting from basic behaviours to more complex emotional expressions.

The studies that tackle this Research Questions (RQs) can be found in Chapter 4. Once the robot is capable of displaying a particular social cue, it is possible to use it during spontaneous HRIs. This can be done via instructing the robot to intentionally or purposefully use a social cue with a goal in mind. This problem can be phrased in the following RQ:

RQ2 How can robots purposefully use social cues in spontaneous HRI?

With the wording *purposefully* in **RQ2** we intend the ability of the robot to intentionally adopt a behaviour in order to reach a well-defined goal. This concept is included in the Perceptual Control Theory (PCT) that shows how our behaviours can be explained as driven by defined goals [120]. We dissect **RQ2** so to focus separately on how interactions can start and how they can be maintained.

Coherently, the following subquestions are defined:

RQ2.1 To what extent, if any, do non-verbal behaviours influence the start of interactions?

This question is tackled with two studies in Chapter 5. First, we explore how a robot can use the social space to start a social interaction. Second, we build a scenario in which the robot uses non-verbal behaviours in a bartending scenario to start a social interaction.

After the focus on the "initiate" transition, we proceed investigating how interactions can be maintained by the robot. Hence, we identified the following subquestions:

RQ2.2 To what extent, if any, do different robot's communication styles maintain interactions?

RQ2.3 To what extent, if any, do different robot's emotional-adaptive behaviours maintain interactions?

The question **RQ2.2** is tackled in a user study conducted *in-the-wild* in which a humanoid robot engaged participants in a quiz game and uses different communication styles during the interaction.

Regarding **RQ2.3** an additional user study is reported in which we investigate if social robots may benefit from employing Emotional Intelligence (EI) when engaged in a conversation with a person. These two RQs are presented in Chapter 6.

The underlying assumption made thus far is that the robot is able to measure interactions. To evaluate its quality or analyse its temporal progression. A metric associated with these aspects is the engagement. It is a dynamic and context-dependent metric that collects real-time data on the interaction and can be used by robots' to dynamically adjust their behaviours. Various operationalisation of engagement are available in the literature, and are usually linked to the defined interaction context. The next overarching RQ is defined as:

RQ3 How can engagement be measured in HRI?

Let us imagine a scenario in which a robot approaches a human in a public space and their interpersonal distance decreases and sets to a value coherent with social interactions [71]. Will this sort of scenario *always* lead to an interaction? Which factors shall the robot consider starting an interaction?

This sort of scenario is expected to happen with increasing frequency, and literature already suggests as central the role of anthropomorphism when starting interactions [171]. For instance, we can concentrate at first with the most prevalent communication mode: non-verbal behaviours. The advantage of considering non-verbal behaviours to assess the interaction is exploited to define a novel engagement metric that assess how HRIs start. Therefore, the following subquestions are defined:

RQ3.1 How to model and measure engagement in case of non-verbal behaviours?

RQ3.2 To what extent, if any, gaze and proximity affect engagement?

Engagement is modelled as a modular and lightweight metric that depends on non-verbal behaviours of the interacting pair, e.g., a human and a robot.

A validation of this metric is done by comparing it with one of the most acknowledged engagement metric [49]. This latter is obtained by training an Artificial Intelligence (AI) model on a prominent dataset (UE-HRI by [23]) with timely annotated labels regarding the interactions.

Deploying a dataset that combines data from annotators with data generated by a robot in a time-sensitive way might suffer from uncontrolled synchronization errors. These, can severely impact the reliability of the dataset and hinder metrics that assumed a precise synchronization. Without reliable data, the advancement of research in HRI risks being undermined.

To address this challenge, a tool for assessing the reliability of datasets with timely annotated subjective measures is developed and tested on UE-HRI [23]. The tool is also presented as part of this thesis. Covering these topics, Chapter 7 presents a novel engagement metric focused on starting interactions via considering non-verbal cues only, and a tool for assessing the reliability of datasets in HRI. The parameters of the engagement metric are learned by solving an optimization problem on a reliable (according to our tool) subset of the dataset UE-HRI [23].

To conclude, this thesis lays the groundwork for understanding how robots can integrate into human social environments through the use of non-verbal cues and adaptive behaviours.

1.5 Structure

The following chapters will tackle the defined RQs, exploring empirical studies that aim to enhance the robot's capacity to start and maintain interactions in real-world settings, as well as the robot ability of assessing engagement. The thesis is organized as follows.

Chapter 2 provides an overview of scientific contribution that are related with the goal of this thesis. It splits the section in a way that is coherent with the structure of the following chapters.

Chapter 3 presents the Spontaneous Interaction State Machine (SISM). It is a model in which distinct states and their relation within spontaneous interactions are defined.

Chapter 4 presents two studies in which robots show different social cues to participants, spanning from cues that are closely related to familiar artefacts like vehicles to the representation of affective behaviour in a multimodal way.

Chapter 5 presents three studies in which robots with different designs use social cues for starting interactions. The first study has a focus on the use of social space, while the second one has a focus on the influence of the context to the social interaction. The last study investigates the impact of our emotional state in how we perceive an AMR navigating in our proximity.

Chapter 6 presents two studies in which robots are instructed to use social cues to maintain the interaction. The first study focuses on the communication style of the robot and show its implementation in a playful scenario. The second one tackles how a humanoid robot can use our emotions to adapt its proximity while conversing with us.

Chapter 7 presents an approach for measuring interactions by using the non-verbal social cues and a way for assessing reliability of datasets in HRI.

Chapter 8 summarizes the scientific contribution of the thesis pivoting on the defined RQs, lists possible limitations of this approach, and highlights open challenges that can be addressed in future work.

Chapter 2

Related Works

Recent years have witnessed significant progress in the field of Human-Robot Interaction (HRI), especially in the area of comprehending how humans and robots can interact in dynamic social contexts. To improve the quality of interactions, many studies have looked into the dynamics of interactions, their measurement, and the ubiquity of non-verbal communication. However, despite these efforts, gaps remain in how to effectively integrate social cues in social robots that are required to conduct spontaneous interactions, a challenge that this thesis seeks to address.

Which social cues are relevant for robots? The meta-analysis in Xu *et al.*[210] provides an interesting research direction regarding the power of robots' social cues. In particular, authors suggest to robot developers to employ a hierarchy of robots' social cues. Prioritising the design of facial expressions, eye gaze, and meaningful movements can facilitate users' to perceive robots as intelligent social actors.

On a similar note, [110] posited that it is essential for scholars to differentiate between primary social cues and secondary ones, grounded in users' evolution-based reactions to media technologies. Primary cues, such as the human voice and human shape, are sufficient but not necessary to obtaining social responses from users. In contrast, secondary cues, including text and machine-generated speech, do not serve as either sufficient or necessary conditions for triggering social responses among users. In contrast to secondary cues, primary cues exhibit greater naturalness, power, intuitiveness, and salience in relation to users' perception of socialness.

The incorporation of gaze cues in the development of collaborative robots, or cobots, has been shown to enhance their perceived sociability and likeability [190]. Furthermore, the movement of the cobot designed to replicate breathing motions suggests positive effects on most measures. The authors used the Godspeed questionnaire from Bartneck *et al.*[19] and the nine scales presented by Heerink and colleagues [79].

Let's now imagine a humanoid robot navigating in proximity to a kitchen table that supports a cup. A person observing this scene could deduce the robot's intention to grasp the cup. They could explain the robot's path to being functional when grasping the cup. However, the robot might have computed such a path only as a result of path constraints—e.g., an obstacle to avoid—and have no information about the cup.

This highlights a critical aspect of HRI: humans tend to assign intentionality to robot behaviours, interpreting actions in terms of goals or purposes that the robot may not actually have. In this scenario, the person perceives a purposeful action where none exists, illustrating how humans naturally apply a "theory of mind" to robots. This tendency can lead to misinterpretations, where the inferred purpose (e.g., grasping the cup) does not align with the robot's actual programming or lack of intent.

This discrepancy between perceived and actual robot intent underscores the need for developing effective social cues in robots to better communicate their actual goals—or lack thereof—to human observers. By incorporating clear, deliberate social signals into a robot's movements and interactions, designers can help manage human expectations and reduce ambiguity in shared human-robot environments. These cues could range from simple gestures indicating intent to more complex, expressive behaviours that clarify the robot's purpose or explain constraints, ultimately fostering smoother, more intuitive HRIs.

Social cues are crucial for robots to join our social environment. However, it is not yet clear how they can be modelled and how robots with various designs can implement them. Considering social robots that will populate our social environments, it is not uniquely defined how these will display and use social cues during their activities in our daily lives.

2.1 On Modelling Social Cues

A robot able to display and communicate its internal state can greatly improve the way it is perceived. The following sections suggest possible ways of displaying social cues, such as using simple and intuitive cues such as lights and sounds or more complex ones like emotions.

2.1.1 Simple Social Cues: Lights and Sounds

The increasing deployment of Autonomous Mobile Robots (AMRs) in everyday environments has motivated research into designing legible robot behaviours that can communicate navigational intentions effectively. Breazeal *et al.*[29] emphasised that non-verbal communication plays a crucial role in making a robot's internal state legible to humans, facilitating smoother interactions. Similarly, Cha *et al.*[37] noted that non-verbal cues in service robots should align with users' expectations in social contexts to enhance understanding and predictability.

To create simple and intuitive communication signals, many approaches utilise lights and sound cues to represent a robot's internal state. Light, for example, is often employed to signal information through variations in intensity, colour, or frequency, while sound has been shown to be effective across different cultural and language groups [104]. Jee and colleagues [90] demonstrated that auditory signals not only conveyed a robot's intentions but also expressed emotions effectively.

Fernandez *et al.*[55] investigated a mobile robot navigating a hallway and using Light Emitted Diode (LED) strips to signal its intention to pass a human participant. While participants initially struggled to interpret the LED signals, a brief demonstration improved their understanding. Shrestha *et al.*[180] compared the effectiveness of arrow-like displays to flashing lights as motion cues during head-on interactions between a robot and a pedestrian, with turn indicators being rated as more intuitive.

Other research has explored the use of gaze and proxemics. Hart *et al.*[75] utilised a virtual agent's gaze on a mobile robot to coordinate navigation with humans, demonstrating that gaze significantly improved users' understanding of the robot's navigational intentions. However, Fiore *et al.*[56] found that gaze cues were less effective for non-humanoid robots, particularly in influencing perceived social presence.

Watanabe *et al.*[203] introduced a system where a wheelchair projected lights on the floor to communicate navigational intent, improving comfort and understanding for both the passenger and surrounding individuals. While these studies provide valuable insights into non-verbal robot communication, many require explicit demonstrations or depend on the robot's anthropomorphic features to be fully understood. Furthermore, sound is typically used only as an attention-grabbing tool in uncontrolled, real-world environments.

With the goal of studying how social cues can be displayed by a robot, our study (see Section 4.1) investigates whether familiar sounds generated by a mobile robot can clearly communicate its navigational intentions to users, enhancing collision avoidance in shared spaces.

The investigation of non-verbal cues in robotic communication corresponds with wider efforts that focus on improving the clarity of robotic behaviours, especially in non-humanoid robots where conventional anthropomorphic signals are lacking. A possible way of displaying a social cue is via exploiting involuntary cues such as emotions.

2.1.2 Complex Social Cues: Emotions

The future of social robots is strictly related to the capability of these to elicit emotions in humans [41]. On this topic, Beck *et al.*[21] evaluated children's ability to interpret a robot's emotional body language, demonstrating, for instance, the impact of head position on the perception of various body postures.

Löffler *et al.*[109] highlights the importance of empowering social robots with artificial emotions that are effective given by a combination of three low-cost output channels (colour, motion and sound).

Rossi *et al.*[161] showed that children aged 3–8 years perceive the robot's behaviours and the related selected emotional semantic free sounds in terms of different degrees of arousal, valence and dominance: while valence and dominance are clearly perceived by the children, arousal is poorly distinguished.

Fisher *et al.*[58] authors selected 27 papers in HRI with a narrow focus on how, despite the morphologies and functionalities, robots can express emotions in an unambiguous way. Their findings suggest how robot's emotional display is expected to happen at very specific moments within interactions, and how different conventions on emotional expressions are adopted in different cultures.

Nivikova & Watts [139] investigate how the body of a non-humanoid robot can be controlled to support the attribution of specific emotions. When combining various communication modalities such as speech, gestures, and visual feedbacks; the resulting behaviours that can be associated with emotions as emotional social cues [211]. However, it is not yet clear to what extent can this be done. Tackling this gap, the study presented in Section 4.2 explores how a non-humanoid robot can display various social cues intended to support the attribution of specific emotions.

Robots that are capable of displaying social cues via either using simple cues such as lights or sound, or more complicated ones like emotions, can start interaction using these cues.

2.2 On Starting Interactions

Social robots must be aware of their surrounding and capable of starting and conducting social interactions. In several applications, social robots are expected to move in our social environments. The following sections suggest the relevance of developing robots that are capable of purposefully start interaction via either using the social space, the context, or the emotions of the person in its vicinity.

2.2.1 Using Social Space and Gaze

Satake *et al.*[171] developed a model that predicts the walking behaviour of a person in the proximity of the robot, plans a path towards them and finally conveys the intention to start a conversation in a non-verbal fashion. The interpersonal space, relative body pose and mutual gaze [146] can be used to capture a snapshot of the evolution of an HRI. The way these variables develop over time can give us more insights into the dynamics of the interaction. Yet, an orchestrated employment of these in a multimodal fashion is expected to improve smooth HRIs [83, 22].

When an interaction with an anthropomorphic robot is about to start, Kendon's model [93] can be used to define the social robot skills [15] and greetings behaviours [78].

Gaze [2, 185], proxemics [137] and body movements [170] are among the non-verbal cues that can be interpreted as social signals and can be used to convey the robot's intentions. Gaze can be manipulated to convey positive or negative robot's mental states and intentions during an interaction [2]. Yet, a robot that stares a human during a social interaction is not positively perceived [214].

The approaching phase in a social HRI provides a first impression that can be used to deduct social intentions. Research highlights that proxemics and the robot's body motions in this phase are pivotal for the users' perception of the robot's intention [83, 137, 164]. The way a robot approaches a human can be interpreted by the latter in different ways [162]. A fast movement towards the human might elicit fear and discomfort [113]. On the other hand, if the robot approaches the human too slowly, the latter might not understand its intention to interact.

Section 5.1 reports a study that builds upon this rationale and explores how a humanoid robot approaching a standing human can convey social intentions via the sole use of non-verbal communication. In contrast to the design choices of [171], we decided to constrain the movement of the humans and focuses on how various non-verbal cues of the robot can influence the perceived intention to interact only during the approaching phase. Beside manipulating social space and gaze, an alternative way of starting interactions can be found in the change of context that might lead to social interactions.

2.2.2 Using the Context

The hospitality sector offer great opportunities to test and develop robotic solutions that can simultaneously address the technical and social challenge. For instance, a robot acting as a bartender might need to prepare drinks for customers and, when left alone, perform tasks such as tidying up or cleaning the working areas.

Ngo *et al.*[138] describe the concept of a robot that can serve many users with only one cocktail option. Their solution is limited in social skills and is designed as an autonomous machine that can repeatedly render drinks.

Foster *et al.*[61] investigate the capabilities of dealing with multiple customers of an anthropomorphic robot acting as a bartender. The robot system tackled a dynamic, multi-party social setting and incorporates state-of-the-art components for computer vision, linguistic processing, state management, high-level reasoning, and robot control.

Authors in [160] developed an autonomous robotic system capable of working as a bartender and interacting naturally with customers. They considered three interactive interfaces (i.e., a totem, a bartending robot, a waiter robot) acting as a centralised system designed around costumers' needs, preferences, and mood. The system consists of two Kuka LBR iiwa 14 R820[1] robotic arms, each featuring 7 Degrees of Freedom (DoF) and equipped with a gripper, which are affixed to a stationary torso. The robot is additionally outfitted with a Furhat robot[2] (3 DoF), featuring a human-like mask that facilitates natural and intricate facial expressions, encompassing realistic speaking movements and expression of emotions.

The strength of the robot described in [160] can be found in its social skills and the ability to render different cocktail requests to different users. The robot uses speech to interact with users, and most of its non-verbal behaviours are provided by the facial gestures of the head.

In contrast to these works, we focus on how interactions can be triggered by a humanoid robot in an environment that includes social and asocial behaviours via using a combination of non-verbal cues and contextual information. The findings of this study are reported in Section 5.2.

Regardless of the context, the dynamics of interpersonal distance are critical when robots navigate in our vicinity. Leaving room to investigate to what extent shall robots modulate their navigational path as a function of our emotions in the moments that can precede an interaction.

2.2.3 Adapting to Emotions

Robots deployed in public spaces can establish interactions by displaying interest in their performed motions [11], emotions [200] and behaviours [201]. Inspired by how humans naturally generate social motions in space, Wen *et al.*[206] explored Inverted Optimal Control (IOC) methods to generate robot motions. Participants found these trajectories to be more appropriate than the control ones.

Yet, appropriateness also depends on where, with respect to us, the

[1] kuka.com

[2] furhatrobotics.com/

robot is navigating. Results from Neggers *et al.*[136] show an asymmetry between participants' comfort when the robot is passing in front of them or behind them.

Lam *et al.*[102] proposed a set of rules for socially acceptable navigation. The rules consider not only the final goal and obstacles on its path, but also whether it should interfere with a human's personal space or another robot's working space.

Zhang *et al.*[213] investigated to what extent a companion robot could track and follow humans at a comfortable distance, while Bera *et al.*[25] proposed a system that combines people's facial expressions and trajectories to enable socially-aware robot navigation.

Ko *et al.*[98] and Raggioli *et al.* [150] tried to understand human intentions (posture and position) and have the robot respond accordingly. In their work, the robot 1) detected the user's behaviour, 2) selected a predefined behaviour based on a Human-Human Interaction (HHI) dataset, and 3) adapted its behaviour based on the user's posture and position. Moreover, in Raggioli *et al.*[150] the human's discomfort is also considered. Similarly, Narayanan *et al.*[134] predicted human emotions by tracking their walking gaits with an onboard robot camera. These predicted emotions were then utilised for emotion-guided navigation, considering both social and proxemic constraints.

Samarakoon *et al.*[169] proposed a novel method for adapting the termination position of a mobile robot approaching a user based on their behaviour and feedback. The authors analysed the skeletal joint movements of the user to assess their physical behaviour. They then fed this information into a fuzzy neural network to determine the appropriate interpersonal distance. The study's findings indicate that users are more satisfied when the robot considers their preferences in its proxemics behaviour.

The work of Papadakis *et al.*[144] analysed HHI and introduced a social map where individuals' personal and social spaces are taken into consideration for human-aware navigation, while in [143], they improved the way to describe the social zones.

Kim *et al.*[95] focused on the importance of social distance in HRI and its relation to the interaction role (supervisor vs. subordinate). In their work, participants identified the comfortable interpersonal distance between themselves and the robot during a task. Depending on the role of the robot, participants had different preferences on the comfortable distance between them and the robot (close or distant). Torta *et al.*[192], explored how a robot should approach a seated participant from different directions and angles. With a questionnaire, the participant could determine a comfortable distance.

Previous research has mostly focused on detecting and computing proxemics and personal space from a purely spatial perspective. In contrast, the work reported in Section 5.3 explicitly considers how emotions influence our perception of robot proxemic behaviours in the moments that can precede an interaction. By addressing this gap, we hope to provide a better understanding of the complex dynamics between humans and robots, paving the way for more effective and meaningful interactions in the future.

2.3 On Maintaining Interactions

Given the ability of a social robot to start an interaction purposefully, can it also maintain it? This question depends heavily on the interaction context. For this reason, the following sections highlights the need for robots to maintain interaction during verbal interactions.

2.3.1 Styling the Communication

A robot capable of changing the style and modes of its communication can persuade humans to provide information or changing their behaviour.

Liu *et al.*[107] provided an extensive review of persuasive robotics and summarised findings on the interaction effects of multiple factors for the persuasiveness of social robots. Saunderson *et al.*[172] presented how a robot's emotional or logical persuasive strategy influences people's decision-making during a game. Their results showed emotional persuasion as the higher persuasive influence strategy, and this might be due to the criticality of emotions in people's decision-making processes.

Ghazali *et al.*[66] evaluated reactance and compliance to persuasive attempts of an artificial social agent, a social robot with minimal social cues, and a social robot with enhanced social cues.

Ham *et al.*[73] conducted a user study that investigates non-verbal persuasive strategies by manipulating the gaze and gestures of a robot narrating a story to the participants. They showed that the robot employing a combination of gaze and gesture incremented robot persuasiveness w.r.t. the robot implementing gesture behaviour only. Similarly to their work, here we also use gaze (in terms of face pose) and gestures to evaluate the impact of the robot communication style on users' attitude to comply with its requests.

Hashemian *et al.*[76] investigated the persuasion capabilities of a robot employing multi-modal interaction on users' free choice of coffee. The authors manipulated the social power of the robot, specifically through the manipulation of its social reward (humorous robot) and expertise (well-informed robot). They found that participants did perceive the robot communication style as significantly different along the persuasive dimensions.

Another interesting approach is presented by Lee *et al.*[105], in which the *foot-in-the-door* technique is implemented in a robot as a persuasive strategy. This technique consists of the robot asking a small request first and then following up with a larger, actual target request. Their results indicated that this strategy could increase the persuasive power of the robot.

In our study, reported in Section 6.1 and conducted *in-the-wild*, we do not manipulate the embodiment of the social agent but the style of its communication with users. The study shows how various multimodal communication styles can be perceived by a person engaged in a quiz game with a humanoid robot.

Similarly to [105, 76], we also underline the importance of multi-modal interaction in building effective behaviour based on persuasion, however, we assess the impact of the robot's communication style also with respect to the personality traits of the users.

With respect to verbal and non-verbal cues, an interesting work is presented by Chidambaram *et al.*[40]. Chidambaram and colleagues conducted a two-by-two experimental study in which four different conditions were designed including verbal and non-verbal communication, namely: no non-verbal cues, vocal cues only, bodily cues only, and both bodily and vocal cues. Their work highlighted the importance of non-verbal cues for improving people's compliance.

Rea *et al.*[151] evaluated the benefits and tradeoffs of various politeness levels for a robot verbally assisting a user in performing physical exercises. Their results showed that participants that interacted with the impolite robot performed more physical activity w.r.t the ones that experienced the polite one.

Green *et al.*[69] implemented six types of verbal persuasion techniques namely commitment, scarcity, concreteness, social identity, emotion and no persuasion. Their work revealed that the content of a conversation with a robot employing a commitment narrative was the most successful with 75% in persuading the users into completing a hidden task.

In contrast to [151], we evaluate the performance of the task based on a quiz, rather than on physical activity. Similarly to [69], we also shape the content of the conversation through the robot's communication style.

Overall, while it is clear that robots' multi-modal interactions have an impact on user behaviours, it is less evident how communication styles can be modelled into robots. Furthermore, most of the works in the field are either conducted in lab settings, resort to convenience samples and are hard to be replicated by other peers.

Styling communication in HRI often involves not only the use of verbal cues but also a sophisticated understanding of non-verbal signals, such as body language, gaze, and proxemics [133]. Proxemics, or the regulation of interpersonal distance, plays a crucial role in shaping the interaction dynamics and creating a more comfortable and engaging experience for users. In particular, the adaptation of these non-verbal cues to the emotional states of humans is critical for maintaining interactions. This brings us to investigate how emotions can be used by the robot to maintain interactions like conversations.

2.3.2 Adapting to Emotions in Conversations

Consider a scenario where a robot is interacting with a person who exhibits signs of anxiety or discomfort. This can be caused by the robot invading their personal space. In such cases, the robot could autonomously adjust its proximity by increasing the interpersonal distance, which helps create a more comfortable interaction space. This adjustment could not only mitigate the individual's stress but also fosters a more effective and relaxed communication environment, supporting smoother and more positive interactions between them. This sort of behaviours could be built by developing social robots with human-like Emotional Intelligence (EI).

Mumm and Mutlu *et al.*[133] investigated the applicability of social-scientific theories from HHI to the field of HRI. They conducted a user study with a social robot and manipulated participants' liking of the robot and the robot's gaze behaviour. The study found that participants who disliked the robot maintained a greater physical distance from it when the robot's gaze behaviour increased. Meanwhile, participants who held a positive view of the robot did not exhibit any variation in distancing from the robot across different gaze conditions. Their results support that the *compensation model* [13] of interpersonal distance can enable robots to conduct more comfortable interactions.

When two humans are talking to each other, their distance is regulated by a constellation of factors such as context, topic, and personal relation [44]. In HRI, changes in proxemics preferences were investigated as related to the pose (e.g., sitting or standing) [164] or the activities the users were performing (e.g., relaxing or working, etc.) [150]. However, these studies did not consider proxemics during verbal interaction with people. In light of the increasing prevalence of robots that are capable of navigating their surroundings and engaging in conversations, it becomes imperative to consider the implications of their proxemics behaviours.

Emotion recognition is crucial for a robot to choose the appropriate behaviour in a given situation. In Castellano *et al.*[34], the authors used the robot Pepper as a Socially Assistive Robot (SAR) and instrumented it with a Facial Expression Recognition (FER) model specialised in detecting emotions in elderly faces. Their results suggest that this type of robot can be accepted as an effective social actor in their cognitive therapy group.

Petrak *et al.*[147] designed an online study to investigate which robot proxemic behaviour (approaching, not moving, moving away) was more appropriate based on participants' expressed emotional state. Their findings suggest that *moving away* is considered inappropriate in most cases. When participants expressed fear, sadness or joy the preferred behaviour was robot *approaching*.

While the studies discussed above provide valuable insights into proxemic behaviours and emotion recognition in HRI, they largely focus on static or structured interactions where robots adjust their behaviour based on limited emotional feedback or gaze behaviour. These investigations do not explore how we perceive fully autonomous robots' adjustments in physical distance during real-time interactions, especially when emotions vary or intensify during the course of a conversation.

This gap in the literature calls for a more nuanced understanding of robot's emotional-adaptive proxemics behaviours. Specifically, whether robots can adapt their interpersonal distance based on moment-to-moment emotional feedback—such as discomfort or anxiety—is not fully addressed.

The study reported in Section 6.2 tackles this challenge by investigating how robots can autonomously adjust their proximity to humans in response to emotional cues, ensuring that interactions remain comfortable and respectful of personal space.

This study emphasizes how real-time emotional feedback can guide proxemic adjustments in conversational settings to enhance interaction quality. Considering how spontaneous HRIs might evolve, a way of measuring interactions as they unfold is needed.

2.4 Challenges in Measuring Interactions

What does it mean to measure a social interaction in HRI? Several authors have tackled this question using a multitude of robot platforms, tailored scenarios and well-defined interactions. Metrics that can be informative in this regard can be extracted from available data logs from real interactions. These can be published in terms of datasets that other researches can exploit. As such, essential for the quality of the interaction is to consider such data source as reliable.

2.4.1 Towards Reliable Datasets

When studying HRI, the quality and reliability of data play a critical role in understanding and interpreting the interaction dynamics. High-quality datasets allow researchers to explore various aspects of behaviour, communication, and engagement in HRI, providing a foundation for future work in the field. These datasets often contain temporal information that captures the nuanced behaviours of both humans and robots, and are frequently logged using state-of-the-art tools like rosbags, which help preserve the temporal sequence of events.

However, while the HRI community has made significant progress in generating and sharing these datasets, a key challenge remains: assessing the reliability of such datasets. Without this, the ability to draw meaningful insights from the data is compromised, potentially leading to misinterpretations or incomplete analyses.

Standardising how results are published may boost the progress of the field [70], as more researchers worldwide seek to replicate studies and benchmark their solutions on existing datasets, ignoring potential quality issues.

Wienke *et al.*[207] proposed a framework for the acquisition of multimodal HRI datasets. Their framework accounted for objective as well as subjective measures, however, the adoption of this approach is limited and requires integration with the event-based middleware named Robotics Service Bus (RSB) [208].

Lazzeri *et al.*[103] developed a platform named HIPOP (Human Interaction Pervasive Observation Platform), designed for multimodal acquisition. HIPOP is a flexible system comprising diverse hardware and software components, enabling the configuration of personalized data collection setups for studies in HRI. By employing modules for capturing physiological signals, eye movements, video, and audio, the platform facilitates comprehensive analysis of both affective and behavioural aspects. Additionally, it allows for the integration of new hardware devices into the existing setup.

A step towards dataset reliability is presented in the Vernissage dataset [89] in which authors implement a post-processing mechanism to validate the synchronicity of all recorded data recorded with an RSB system.

Despite these works attempting to standardise how datasets are built in HRI while overcoming platform-tailored acquisition strategies, their adoption is still limited. This result is the product of choosing the rarely used RSB middleware while focusing on high-level data types such as physiological signals. To counter this, Section 7.1 show how the widely adopted middleware Robot Operating System (ROS)[3] can be used to build datasets, thanks to its own logging mechanisms, with a focus on low-level data types that can be recorded during an HRI. A tool for assessing the reliability of

[3] https://www.ros.org/

datasets made in ROS is implemented and tested on the popular UE-HRI dataset [23].

These metrics not only enhance our understanding of user experiences but also serve as critical benchmarks for improving the design and functionality of social robots. By integrating robust engagement metrics with reliable datasets, researchers can ensure a comprehensive approach to evaluating HRI, ultimately leading to more effective and user-centred robotic systems.

2.4.2 Measuring Interactions

Engagement metrics provide valuable insights into the level of interaction between humans and robots, allowing researchers to evaluate the success of various interaction strategies. Recent advancements in the field have led to the development of sophisticated metrics that assess engagement based on both behavioural and emotional dimensions. For instance, various approaches have been proposed to quantify engagement by analysing user responses, robot behaviours, and contextual factors during interactions. Before providing how engagement is assessed in the literature, it is important to have a clear overview of its definition.

Sidner *et al.*[181] define engagement as “the process by which two (or more) participants establish, maintain, and end their perceived connection. This process includes initial contact, negotiating a collaboration, checking that the other is still taking part in the interaction, evaluating whether to stay involved, and deciding when to end the connection.” In [27], the authors define it as “the process subsuming the joint, coordinated activities by which participants initiate, maintain, join, abandon, suspend, resume, or terminate an interaction.” This interpretation builds upon the one in [181], by considering the concepts of abandon, suspension, and resuming the interaction experience.

This reflects in modelling engagement as a continuous and synchronous process that clearly has a beginning and an end [52]. Poggi [149] defines engagement as “the value that a participant in an interaction attributes to the goal of being together with the other participant(s) and of continuing the interaction”. Hence, it models engagement as a quality metric of the interaction.

O’Brien & Toms [140] define it as “the quality of user experience characterized by attributes of challenge, positive affect, endurability, aesthetic and sensory appeal, attention, feedback, variety/novelty, interactivity, and perceived user control". They model is as four discrete events that explain the engagement dynamics namely: point of engagement, period of sustained engagement, disengagement, and re-engagement.

With a narrower focus on the social dimension of engagement, Salam & Chetouani [166] describes it as the "measure of the intention-to and the quality-of interaction as perceived by the user"

On a different approach, Moschina *et al.*[132] described social engagement as "a core social activity that refers to an individual's behaviour within a social group." The overview provided by Oertel *et al.*[141] highlights the complexity of *engagement* regarding Human-Agent Interaction and while

The systematic review from Sorrentino *et al.*[182] focuses on the conceptualization and automatic detection of engagement in HRI, examining various methodologies and their implications for developing socially intelligent robots. It synthesizes existing literature, identifies research barriers, and outlines future challenges in understanding and assessing engagement during interactions between humans and robots.

Several authors have operationalised engagement and instrumented robots with it. For example, the system proposed in Del Duchetto *et al.*[49] allows estimating engagement in real-time, given that the user and the robot are already interacting. In their work, the authors used three independent annotators to classify video clips of users interacting with a robot and trained a Convolutional Neural Network (CNN), a Long-Short Term Memory (LSTM) network with the annotated data. Love *et al.*[111] presented a two-layered proactive system that extracts high-level social features from low-level perceptions and uses these to decide whether starting or maintaining HRIs.

Lemaignan *et al.*[106] presented a rule-based metric *with-me-ness* to allow a robot to assess in real-time the focus of attention of the interactants. This metric, priorly introduced in Sharma *et al.*[177], can evaluate how well the user's facial pose is aligned with an expected pose during a task with a robot.

Youssef *et al.*[212] investigated deep learning techniques to detect how user engagement decrease in real-time using a dataset of spontaneous interactions with a humanoid robot.

Kesim *et al.*[94] presented a dataset for multimodal engagement with a humanoid robot head using the classical operationalisation of the metric given by [155], where the mean time between successive "connection events" is used for assessing engagement. The same work hypothesizes a minimum occurrence frequency between these events as the process mechanism for maintaining engagement.

Nasir *et al.*[135] investigated engagement in learning scenarios and define *Productive Engagement* as the "level of engagement that maximises learning" via considering both task and social engagement, and verbal and non-verbal communication channels. The same work suggests that for maximizing learning outcomes the robot shall seek to optimize the engagement rather than maximize it.

With the aid of electrophysiological data (EEG signals), Ehrlich *et al.*[53] investigated if an established gaze contact can convey the roles of initiator or responder of the interaction. For this purpose, the authors trained a Support Vector Machine (SVM) for offline classification.

Another cue that can greatly influence interactions, besides gaze, is proximity to a robot. Kim [95] show that interactions with a robot can initiate given an appropriate interpersonal distance. However, proximity is a social feature that can vary with individuals and context [126].

Engagement is also a function of the dual system formed by the user and the robot, and according to the intention of each party, the interaction can start, progress or finish. In this direction, Ivaldi *et al.*[87] presented a two-phase interaction study, in which the authors showed that the starting role of the robot in the first phase of the task has a consequence on the rhythm of the interaction in the second phase. Their work showed that the behaviour adopted by the robot influences the user response.

Anzalone *et al.*[12] derive engagement from the mutual communication established between the human and the robot. The authors performed static and dynamic analysis of the body motions and tackled the metric as a possible response (or reaction) given the robot behaviours.

Webb *et al.*[205] developed a game with a cross-platform game engine that allows the simulation of social interactions with virtual characters and introduced the metric named "visual social engagement" as a value that simultaneously depends on the distance between two agents and their mutual gaze. While playing the game, the user can control the behaviour of one virtual character and their game objective involved approaching other characters and interacting with them. The authors defined the metric as symmetric w.r.t the social agents. However, this study contends that considering the intentionality of each social agent's participation in the interaction can result in a more robust model of engagement. This departure from a symmetrical framework enables a more nuanced understanding of the metric and the complex relationship among social agents' intentions. A similar approach was already proposed by [72] and underlined recently by Maniscalco *et al.*[116] where authors highlight the importance of considering bidirectional communication when assessing engagement.

Future engagement frameworks urge for defining engagement metrics that exploit the context of the interaction [182]. The use of black box systems for assessing engagement like Del Duchetto *et al.*[49] involves the risk of not knowing which environmental variable is truly influences engagement. A step towards explainable engagement metrics is done in Love *et al.*[111], however their work relies on the engagement operationalisation of Webb *et al.*[205]. This latter models engagement as a shared contribution of the behaviour of the social agents (see section 3.1 in [205]).

Yet, the definitions of engagement provided at the beginning of this section show that 1) it is a value related with the interaction-context and 2) it is a perceived phenomenon. Meaning that there *might* be an asymmetry in how each social agent is perceiving engagement. This approach follows the rationale of [87] in which interactants are entangled in the interaction bidirectionally.

Tackling this gap, Section 7.2 presents a novel engagement metric that relies on explainable social features and highlights its asymmetry.

Chapter 3

Towards Spontaneous Interactions

Developing robots with human-like aspects may induce people to interact using familiar social norms [173] and reduce the need for learning new interaction interfaces. This idea goes beyond just making robots that look or speak like us; instead, it entails developing artefacts that are able to recognise, decipher, and react to social cues in a manner that is similar to what people commonly do. A possible way of achieving this is by empowering robots with human-like comprehension of what a spontaneous interaction is.

We can anticipate several use-cases where we expect social robots to behave spontaneously. Building on the example of the robot employed in a busy airport presented in Chapter 1, we investigate a possible model for controlling the robot's behaviours purposefully while addressing spontaneous interactions.

3.1 The Approach

This thesis introduces the Spontaneous Interaction State Machine (SISM) as a finite state machine that models spontaneous Human-Robot Interactions (HRIs) and captures their cyclicity. SISM is designed as an operationalization of a *context model* as defined by [129] with a focus on spontaneous interactions. Finite State Machines (FSMs) are commonly used to model systems with a finite number of states and a known logic to transition among them. The idea is to abstract what a *spontaneous* interaction consists of and model a finite number of states and transitions that represent it. These are designed to represent either the presence or absence of an interaction and its occurrence over time.

The model is also inspired by the General Aggression Model (GAM) [5] which takes into account how aggression is influenced by social, cognitive, psychological, developmental, and biological aspects (see Figure 3.1). Such a model stresses the impact of proximate process to explain how circumstances and people affect perceptions, emotions, and arousal, which in turn impact judgement and decision-making processes, which ultimately impact the consequences of aggressive or non-aggressive behaviours. The proximal processes act as learning trials with each cycle, which influences the growth and accessibility of aggressive knowledge structures. With our model for spontaneous interaction in HRI, we stress how a social robot, developed for autonomous interactions, shall store information at every interaction and potentially use it for the next ones.

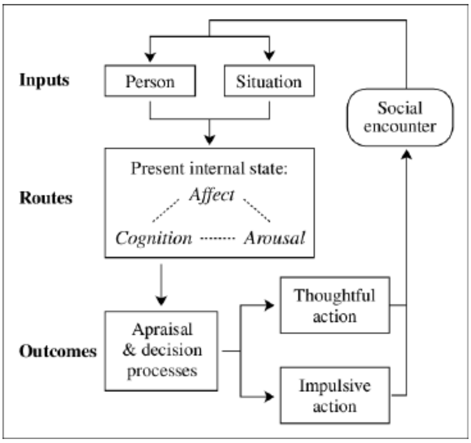


Figure 3.1. Diagram of the General Aggression Model (General Aggression Model (GAM)).

3.2 The Model

The idea is that social robots should continuously assess their contribution to the social environment given a well-known model, e.g., a FSM, of it. The continuous assessment is intended as a *context recognition systems* as defined by [129], while the well-known model is what the same authors refer to as a *context model*. The primary driving force behind the development of this model is the constant fluctuation between presence and absence during autonomous interactions. In other words, if it is not interacting with anyone, it is either starting or ending an interaction.

A possible operationalisation of a *context recognition system* is the level of engagement during an interaction. Another method could involve evaluating the user's emotional state or the firmness of their hands during a handshake, as suggested by [198].

Given the way that robots will be incorporated into our Society, we must find a means of communicating how each interaction affects subsequent ones. Notice that, SISM does not model in any way the aggression in HRI, instead, its main scope is to conceptually separate the presence and absence of social interactions and their occurrence in time. Its representation is available in Figure 3.2.

The state machine is divided into two primary phases:

- *Not Interacting*: This includes the *Pre-Interaction* and *Post-Interaction* states.

- *Interacting*: This encompasses the *Social Interaction* state.

The three self-exclusive states for the model are defined as:

1. **Pre-Interaction State** Here, the robot is in a preparatory state before starting any social interaction. The robot may be gathering environmental or contextual cues to determine the appropriateness of engaging with a person or more. This is where the robot evaluates the social context to decide if interaction should be initiated. From this state, the robot expects to initiate a social interaction. No assumptions are made on the current task performed by the robot, as it can be in an idle state or performing an autonomous task without direct involvement of a person.

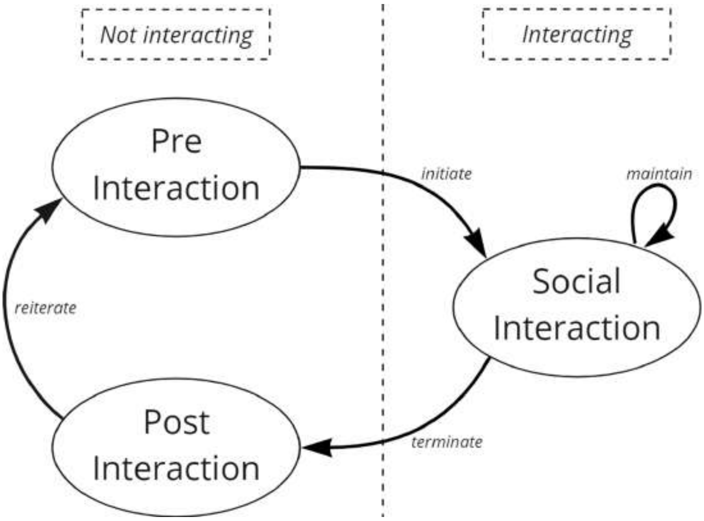


Figure 3.2. Spontaneous Interaction State Machine (Spontaneous Interaction State Machine (SISM))

2. **Social Interaction State** Once the robot has moved to the *Social Interaction* state, it is engaged in active communication or task-based interaction with a person. This state is crucial for maintaining meaningful exchanges and ensuring the interaction remains appropriate, fluid, and effective. In this state: The interaction may be sustained over time through a process of maintenance (e.g., ongoing dialogue, cooperative behaviour). When the interaction reaches its end, the robot transitions to the next state. Yet, the interaction can be controlled by the robot by either *maintaining* or *terminating* it.

3. **Post-Interaction State** After the social exchange is complete, the robot enters the *Post-Interaction* state. Here, the robot may process information gathered from the interaction, storing it for future use or analysing its performance. This stage is essential for enabling the robot to learn from its past interactions, improving its behaviour for the next encounters.

The transitions are designed to create a loop where the robot is constantly preparing for, engaged in, or terminating an interaction. Notice how this idea of constantly observing the social environment fits particularly well with the *context recognition model* as introduced by [129].

SISM is especially useful in environments where autonomous operations of robots are required, and potential social interactions might occur, such as in public spaces including museums, airports, or hospitals. In such environments, it is imperative for the robot to possess a well-defined framework for determining the appropriate moments to engage, strategies for maintaining interaction, and criteria for concluding or processing the interaction. Through repeatedly going through these states, the robot maintains adaptability and responsiveness to the social dynamics of its environment, thereby potentially enhancing its social acceptability and functionality.

Furthermore, the model emphasises the significance of comprehending the contextual nature of interactions. Social contexts are dynamic, necessitating that robots evaluate nuanced signals such as non-verbal communication, levels of human engagement, or indications of disinterest—prior to starting or concluding interactions. The state machine facilitates a distinct, cyclical process for this purpose, improving robots' behavioural flexibility and their ability to dynamically adapting to the social environments in which they operate. The link between GAM and SISM can be noticed with the feedback shown by the "Social Encounter" in GAM that is modelled as the *reiterate* transition in the SISM. The rationales can overlap. After every proximal encounter or *Social Interaction* the robot could learn and store valuable information that can be used for future interactions.

Moving from one state to another, is done by the defined transitions. These can be driven by metrics that are available to the robot at any given time. This characteristic of the model allows context-dependent operationalisation of the transitions. This kind of metric tracks how social interactions unfold around the robot. Measuring social interactions is a hot topic in HRI, yet there is hardly any consensus on what we actually mean when we talk about measuring one. On this topic, the keyword engagement is widely used and several authors tailor it to their use-cases.

Following the work of Menenez *et al.*[129] we proceed in studying engagement as an operationalisation of a *context recognition model* when considering spontaneous HRIs. Emphasising that different interaction context shall use different engagement metrics.

We also highlight a few limitations of SISM. First, the need for selecting an operationalisation of engagement that fits the interaction context. Second, the logic for executing a transition is still defined by a threshold or a known logic on the given metric. For instance, if the level of the metric exceeds a known threshold, it translates to the interaction about to initiate. Defining this sort of logic is surely an open question but falls outside the scope of SISM.

Overall, SISM poses itself as an abstract representation of a spontaneous HRI, but does not assume any communication channel for the transitions. The reason for this is that interactions can be triggered with several communication channels [178, 179, 171]. For instance, one could start a conversation via simply gazing at another person or gently tapping on their shoulder.

Chapter 4

Showing Social Cues

Less is more.

Ludwig Mies van der Rohe

We intentionally craft social robots to leverage individuals' key social and relational abilities. Designers have leveraged the inherent human inclination to attribute human characteristics to non-human entities. For instance, we often perceive faces in arbitrary arrangements of objects or shapes, a phenomenon referred to as *pareidolia*, and are inclined to interpret social implications in the movements of geometric figures, as shown by Heider and Simmel [80].

Building robots for social environments requires careful consideration of how nearby humans will perceive them. Regardless of the robot's design, communication capabilities, or the task at hand, this chapter examines how robots can exhibit social cues through either simple, familiar indicators (**RQ1.1**) or more intricate and advanced expressions such as emotions (**RQ1.2**). To what extent is the robot able to communicate intentions? Which communication interfaces can it use? Targeting these questions, this chapter focuses on the following publications:

> Georgios Angelopoulos*, Francesco Vigni*, Alessandra Rossi, Giuseppina Russo, Mario Turco, and Silvia Rossi. Familiar acoustic cues for legible service robots. In 2022 31st IEEE International Conference

* co-first authorship

on Robot and Human Interactive Communication (RO-MAN), pages 1187–1192. IEEE, 2022.

Francesco Vigni, Alessandra Rossi, Linda Miccio, and Silvia Rossi. On the emotional transparency of a non-humanoid social robot. In International Conference on Social Robotics, pages 290–299. Springer, 2022.

4.1 Showing Familiar Cues

When navigating in a shared environment, the extent to which robots are able to effectively use signals to coordinate with human behaviours can increase social acceptance. This section discusses the results of a study that investigates whether familiar acoustic signals can improve the legibility of a robot's navigation behaviour [11].

The problem lies in the fact that a robot, even if it doesn't want to interact with humans, must be able to perceive its intentions when moving in an environment where humans are present. We believe that a critical aspect is that users need to be familiar with the cues used by a robot to communicate its intention without explicitly training people to read them. Commercial passenger vehicles, for instance, have turning sounds and lights to indicate when to change the direction of travel. Many vehicles can also rely on an intermittent sound that changes its frequency to signal an approaching obstacle. The world unambiguously recognises these simple communication modalities, which have been around for many years.

In this regard, in this section we present a study that uses familiar cues that are easily correlated with regular vehicle cues, such as blinking lights for signalling turns or intermittent sounds related to proximal obstacles, on a standard Autonomous Mobile Robot (AMR). We decided to investigate whether it is possible to seamlessly transfer the semantic knowledge from vehicles to mobile robots. In doing so, this study tackles the question "How can robots effectively display familiar, simple social cues?" (see **RQ1.1**).

4.1.1 Methods

An AMR like the iRobot Roomba is instrumented with an *arduino* microcontroller [17] and prototyped Printed Circuit Board (PCB) to control

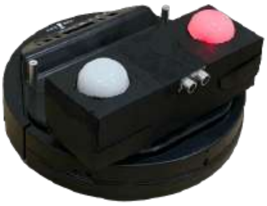


Figure 4.1. Picture of the assembled prototype with mounted blinkers.

small magnetic speakers, a buzzer, a proximity sensor, and two independent Light Emitted Diodes (LEDs) positioned on top of the robot, as shown in Figure 4.1.

The user study was performed by selecting the following experimental conditions as a subset of non-verbal communication modalities used by common vehicles, namely:

- ***Navigational Cue 1*** **(NC1)**: The robot's produced an intermittent tone with a constant frequency, similar to the turning indicator sound of a vehicle. We believe that the use of a turning indicator sound could elicit a directional intention in people.

- ***Navigational Cue 2*** **(NC2)**: The robot used a red LED and a speaker to produce a paired intermittent switch with constant frequency. This cue resulted in a synchronous use of light and sound (i.e., when the light is on, the speaker produces a tone). *NC2* has been designed to provide turning direction (via the right led) while producing the same sound profile presented in *NC1*. In this cue, the sound is not intended for conveying directional intent but to attract attention [100].

- ***Navigational Cue 3*** **(NC3)**: The robot produced an intermittent tone with a frequency inversely proportional to the read of the proximity sensor. This cue aims to mimic the tone produced by a vehicle (e.g., connected to parking sensors) that is approaching an obstacle.

The tone employed frequency variations to convey the remaining distance to the obstacle.

The speaker and the buzzer differ in tone and modulation. The buzzer was used to elicit an interaction that could recall a vehicle during a parking manoeuvre. The buzzer emitted an intermittent tone whose frequency increased as the robot approached an obstacle, i.e., the user. The sound produced by the speaker, instead, was designed to mimic the noise of an active turning light as perceived inside a vehicle. Each LED was installed in an opaque glass-shaped container so that its light could use a higher surface. Each container was located on the top-side of the robot, mimicking the typical lateral position of turning lights in a vehicle. Similar design strategies can already be seen in industrial AMRs with respect to lights [16] and sounds [109]. We designed a between-subject study where different participants tested each condition so that each participant was exposed to an experimental condition only once. With this design, we recruited one hundred twenty participants that allowed to detect an effect size of $d = 0.25$ with 0.90 power at an alpha level of $\alpha = 0.05$. We believed that sound plays an essential role in the communication of intention, however, it conveys a clearer directional motion when combined with the use of the lights. Therefore, our hypotheses were:

- **H1** the navigational cue *NC2* is more legible than *NC1*
- **H2** varying the frequency of the sound in *NC3* will improve the clarity of the robot's intention compared to *NC1*

The video clips of all the conditions had the same shooting angle, duration, environment, and lighting conditions. In particular, the videos were recorded in a long corridor (145cm wide, 320cm long) with the camera positioned at the opposite end, facing the robot. The camera was at a fixed position in all conditions to avoid variations in the field of view. The video clips lasted 10 seconds, and during the first 5 seconds, the robot moved in a straight line starting at the beginning of the corridor towards the camera. Then, the robot approached while signalling using navigational cues as per the experimental conditions. The robot used only a navigational cue but did not complete the next movement (i.e., turning or moving forward) to hide the navigational goal of the robot, since we intended to evaluate participants' understanding of the robot's intent. The online study was distributed to participants via social media and to the University's community members. To evaluate people's perceptions and to understand the legibility of the designated cues, a brief post-interaction survey comprising 5-point Likert and nine scale questions was provided to the participants. The questions can be clustered as follows: (in brackets are the labels as used in Figure 4.2):

1. Comprehensibility
   - The robot's behaviour was misleading (*Misleading*)
   - I quickly understood the robot's behaviour (*Understandable*)
   - It is difficult to understand what the robot intended to do (*Unclear*)
2. Reliability
   - The robot was deceptive (*Deceptive*)
   - I am weary of the robot (*Draining*)
3. Social compatibility, comfort, and friendliness
   - The robot's behaviour would be socially compatible with a pedestrian's environment (*Socially compatible*)
   - The robot's behaviour made me feel comfortable (*Pleasant*)
   - I liked the robot (*Likeable*)

The sample of 120 participants was distributed for the three navigational cues conditions as follows: forty-three participants in *NC3*, 39 and 38 participants in *NC1* and *NC2*, respectively. An independent samples t-test with 95% confidence intervals was used to determine if a difference exists between the navigational cues. Figure 4.2 shows the average responses grouped by navigational cues.

4.1.2 Results

In the question on *Misleading*, *NC1* scored significantly higher than *NC2* with a mean of 2.77 (0.44 to 1.52), $t(73.958) = 3.636$, $p < 0.01$. There

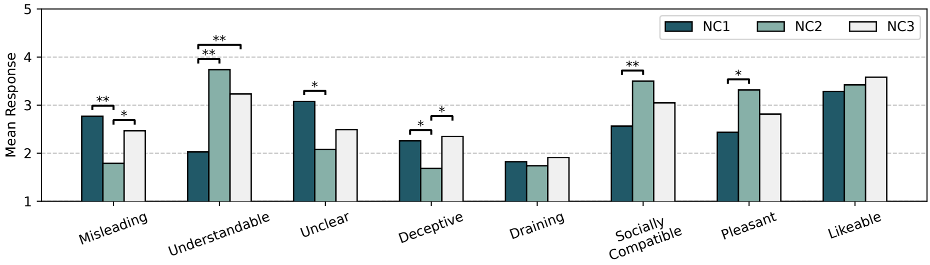
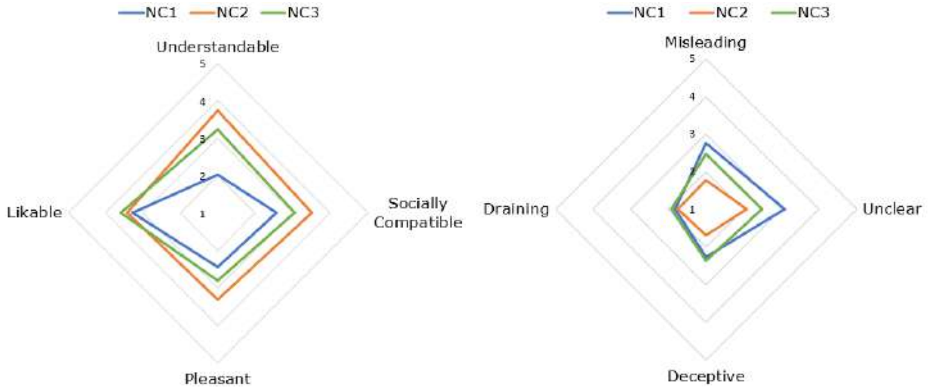


Figure 4.2. Average of responses to the questionnaire, significant differences between navigational cues have been indicated with * for $p < 0.05$ and ** for $p < 0.01$.

was also a statistically significant difference between *NC2* and *NC3*, with *NC3* scoring higher than *NC2*, 2.47 (-1.19 to -0.16), $t(79.000) = -2.604$, $p < 0.01$. Although the averages scored below the mean of 3.00 (on a 5-point Likert scale), users find *NC2* (i.e., the robot using light and sound to convey directional intent) to be the less misleading navigational cue. In the question on *Understandable*, *NC1* scored significantly higher than *NC2* with a mean of 3.74 (-2.25 to -1.18), $t(74.985) = -6.368$, $p < 0.01$. A statistically significant difference between *NC1* and *NC3* was observed, with *NC3* scoring higher than *NC1*, 3.23 (-1.73 to -0.68), $t(79.079) = -4.557$, $p < 0.01$. Hence, varying the sound profile and intermittent frequency (*NC3*) can significantly contribute to conveying directional intent. A significant difference is also found on the *Unclear* scale between *NC1* and *NC2*, with *NC1* scoring higher than *NC2*, 3.08 (0.38 to 1.62), $t(74.723) = 3.214$, $p < 0.01$. These are in line with the results in the *Understandable* scale, and people found clearer a robot that employs light and sound (*NC2*) to communicate directional intent rather than one using only sound (*NC1*). Considering the responses to the question *Deceptive*, *NC1* scored significantly higher than *NC2* with a mean of 2.26 (1.14 to 1.00), $t(72.476) = 2.651$, $p < 0.01$, and a statistically significant difference between *NC2* and *NC3*, with *NC3* scoring higher than *NC2*, 2.35 (-1.14 to -0.19), $t(73.504) = -2.804$, $p < 0.01$. These results show that participants found *NC2* to be the less deceptive navigational cue. Moreover, no significant difference was observed between *NC1* and *NC3* on this scale. Finally, on both questions *Socially compatible* and *Pleasant*, *NC2* scored



(a) Participants' outcomes demonstrated a trend of higher mean ratings in Socially Compatible, Pleasant, and Understandable for the *NC2* and Likeable for the *NC3*.

(b) Participants' outcomes demonstrated a trend of higher mean ratings in Misleading, and Unclear for the *NC1*, Draining and Deceptive for the *NC3*.

Figure 4.3. Positive and Negative effects of the Navigational Cues.

significantly higher than *NC1*, respectively, with a mean of 3.50 (-1.46 to $-0.41, t(74.867) = -3.575$, $p < 0.01$), and 3.32 (-1.42 to -0.35 with $t(74.473) = -3.279$, $p < 0.01$). These outcomes show that the users considered a robot that signals its motions using an acoustic tone synced with a blinking LED (*NC1*) more congruous for a social environment than one using only the acoustic tone (*NC2*).

Considering participants' responses to the *Draining* and *Likeable* questions, we did not observe any significant difference between the three navigational cues.

The questionnaire's entries can also be grouped into *positive* (i.e., Understandable, Socially compatible, Pleasant, and Likeable) and *negative* (i.e., Misleading, Unclear, Deceptive, and Draining) scales (see Figure 4.3). If the values are higher in these questions, it can be interpreted as better a robot's behaviour, while the second might indicate the worst one. The strength of the *NC2* cue is further reflected both in the positive effects and the negative effects of the post-interaction survey. Figures 4.3a and 4.3b provide useful insights into the relative quality of the navigational cues. In particular, we can observe a trend of higher mean ratings for the positive questions in the *NC2*, suggesting it was the preferred experimental condition. It is also interesting to notice that the inverted coaxial order of navigational cues between the negative and positive questions marked *NC2* as the one with higher surface in Figure 4.3a and lower surface in Figure 4.3b.

Finally, these results can be interpreted in light of **RQ1** with the variation that consider non-humanoid robot designs such as AMRs. In particular, this study tackled the subquestion **RQ1.1** that states: "How can robots effectively display familiar, simple social cues?".

Our results highlight a significant legibility improvement when the robot used both light and acoustic signals to communicate its intentions, compared to using only the same acoustic sound. Additionally, our findings highlight that people also perceived differently the robot's intentions when they were expressed through two frequencies of the mere sound. The robot using such cues might swiftly transition from "Pre-Interaction" state to "Social-Interaction", simply by approaching and using cues that humans can understand easily.

Regarding **RQ1.1**, we showed that a possible way for displaying social cues on an AMR is to instrument it with cues familiar to us. The key point is found in using cues with a clear semantics, like blinking lights for signalling turning and frequency-varying beeping tones to signal interpersonal distance. On the one hand, participants were able to understand what the robot was doing (see *Understandable* in Figure 4.2); on the other hand, this robot exploited a non-verbal communication strategy based on familiar cues from common street vehicles.

Limitations We identified two main limitations in this study. Despite it shed some lights on using cues that people are familiar with, it also relies on the subjectivity of *familiarity* of different social cues. Therefore, a procedural and objective way of defining the *familiarity* degree of social cues, is lacking. The post-interaction contains a set of questions to specifically target the hypotheses. We are aware that for measuring legibility, surveys like [10] could also fit, however, following the recommendations of [82] we designed short post-interaction survey to minimize participants' fatigue.

4.2 Showing Emotional Cues

So far, this chapter has described how social robots can exploit cues from other familiar artefacts, like regular passenger vehicles. However, considering Human-Human Interaction (HHI), we strongly rely on our Emotional Intelligence (EI) to interact between each other [31, 122, 127].

People are able to communicate and interpret multimodal communication signals, including natural language, gestures, poses, and body language. In addition to those, they might engage other humans with a bidirectional and mutual understanding [176] that allows them to understand and predict others' intentions and behaviours.

Current literature has identified a number of social cues that could influence people's perception of a robot as a social entity and, as a consequence, their behaviours and trust towards a robot during an interaction [158]. There is uncertainty about whether all cues in a multimodal interaction impact its quality equally, or if some are overshadowed by others [163]. This section presents a study that examines the relationship between emotions and the behaviour of a non-humanoid social robot. In doing so, this study tackles the question, "How can robots display complex social cues, such as emotions?" (**RQ1.2**). The robot used is the *ClassMate* developed in collaboration with the Italian company Protom Robotics[1] and is designed to help the learning experience of students in schools. It is a grounded social robot for classroom environments and allows the development of social expressions in terms of body motion, facial expression, tactile interaction, and sounds [45]. Its design allows for deployment on top of a standard desk without the need to secure the robot to it.

The study explores the multimodal representation of affective behaviours by robots. Furthermore, we are interested in the extent to which interacting users correctly identify these behaviours as emotions. As target emotions, we used the distinctive universal emotions defined by Ekman [54]: anger, fear, disgust, sadness, joy, and surprise. Following the robot's design, we consider facial expression, body motion, and sound as components for achieving multimodal interaction.

[1] protomrobotics.com

4.2.1 Methods

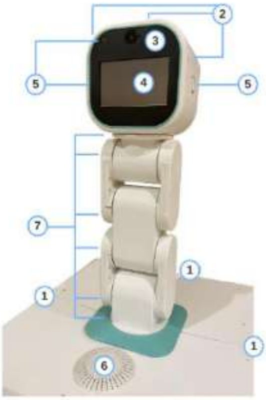


Figure 4.4. *ClassMate* Robot

The *ClassMate* Robot is an open chain robot with 6 Degrees of Freedom (DoF) implemented as revolute joints. The robot could be divided into (fixed) base, body and head. The base allows a rotation of the body along the vertical axis; the body contains 4 parallel-axes joints. Finally, the head is controlled by a revolute joint whose axis is orthogonal to its parent's. Figure 4.4 shows one of the first prototypes of the robot and highlights (1) Infrared (IR) Sensors, (2) Touch Sensors, (3) Camera with a built-in microphone, (4) LCD Display, (5) Left and right RGB LEDs + Frontal camera flash, (6) Sound Sensors and (7) Motors. The robot is designed to engage students, teachers and school personnel in social interactions while providing different functionalities, such as small talks and learning applications [45]. The *ClassMate*'s framework has been developed following four main principles that allow an easy personalisation, update, and extension of the available skills and applications: 1) the robot needs to be interacting and have personalised behaviours, 2) the robot needs to be able to have natural and social interactions, 3) new applications can be easily added by non-programmers, and 4) the applications and services provided need to be perceived as part of the robot and not external tools.

To this purpose, affective modalities can be used by social robots to convey their internal states and intentions [30], and improve the success of

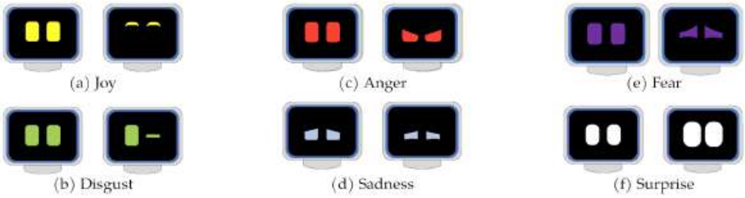


Figure 4.5. Examples of facial expression with relative intended emotion.

the social interaction. The social component of the interaction is manipulated on the *ClassMate* robot's facial gestures, body motions, and sounds, as suggested by [109]. In particular, we endowed the robot with the capability of expressing Ekman's basic emotions (joy, sadness, disgust, fear, surprise, and anger) [54].

The screen located at the end of the chain represents the head of the robot and displays two simple eyes on a black background. The shape and colour of the robot's eye animations have been designed considering relevant studies [148, 46]. Figure 4.5 shows examples of the facial expressions designed in this study paired with the intended emotion.

The body movements of a robot are also used to convey emotions [159]. However, the kinematics of this robot allow limited motion of the joints, so the range of emotional expressions that can be designed is also limited. To convey emotions, we can rotate the last joint (head), control the body to represent "closeness" or "openness" to the interaction [123], and rotate the whole body using the base joint. As discussed in [121], the emotions that a robot's body can express surely depend on its design and anthropomorphism. Here, only three separate body expressions are implemented. Bearing in mind that across all the body motions the robot always starts at the initial configuration (Figure 4.6a).

Figures 4.6b, 4.6c, 4.6d show the final configuration of the body at the end of the expression of each motion. Considering the sound utterances, they can be used to mimic the natural back-channelling cues that are often used by humans to express a specific emotion. In Human-Robot Interaction (HRI), back-channelling cues are important as they can be used to maintain a person engaged with a robot or to attract attention [161].

The purpose of this study was to evaluate which is the minimum type of modalities needed by the *ClassMate* robot for expressing internal states and responses to effectively communicate with people. This is not a trivial task since developed emotions are not always perceived as intended, both in humanoid and non-humanoid robots [159]. A misinterpretation of the emotions may have negative effects on the legibility of the robot's intentions and, as a consequence, on the overall success of the interaction. In this context, we classified several multimodal affective (paraverbal and nonverbal) behaviours according to people's perceived emotions. In particular, we hypothesised that multimodal interactions for such a non-humanoid social robot improve the legibility of its simulated emotions. Therefore, we combined three modalities to identify which are the social cues that make transparent a robot's emotional state for people. We conducted an online questionnaire-based study that was organised as a between-subject experimental design to evaluate the perceived expressions of the robot's animations. Participants watched several animations in which the *ClassMate* robot simulated emotions using: 1) *C1* only one modality, i.e. facial expressions; 2) *C2* facial expressions and body motions; and 3) *C3* facial expressions, body motions, and sounds. Overall, we developed six robot affective behaviours that mimic distinctive universal emotions (anger, fear, disgust, sadness, joy, and surprise). In condition *C1*, the robot's eyes displayed on the screen were white eyes with a fixed shape[2]. In condition *C2*, the robot's eyes assumed the colour[3] as depicted in Figure 4.5. In

[2] The animations used in *C1* condition can be viewed on https://tinyurl.com/2a9bhjzj

[3] The animations used in *C2* condition can be viewed on https://tinyurl.com/

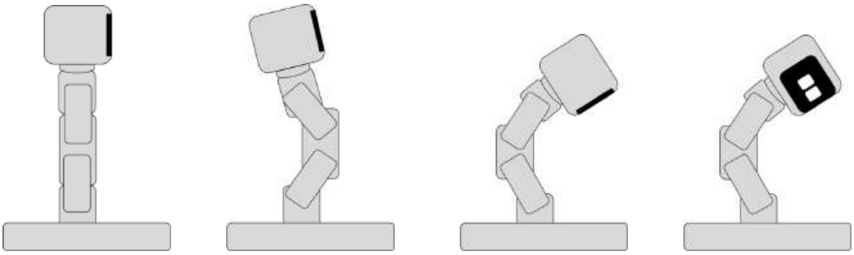

(a) Initial **(b)** Open **(c)** Close **(d)** Close Side

Figure 4.6. Examples of *ClassMate*'s body motions for emotional expression.

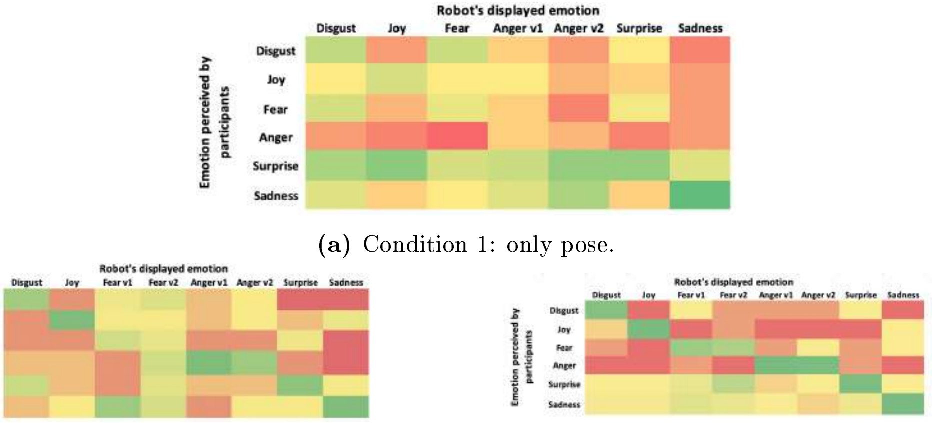
condition *C3*, we used the same facial animations of *C2*, and we added paraverbal sounds that are often used by people to convey the respective emotion[4]. In our design, condition *C3* is the baseline, and we expect that people would clearly associate these animations to the respective emotional expression. The presented animations also included two variations of the fear and anger emotions using different movement directions. The combinations of the body expressions and emotions were inspired by the results presented in Löffler *et al.*[109]. Each animation lasted about 3 seconds and was displayed in random order to the participants. During the different stages of the questionnaire study, we asked participants to complete several questionnaires. A pre-experimental questionnaire collected participants' demographic data (age, gender) and their previous experience with robots, including what kind of previous interactions and types of robots they interacted with. After each video, participants were asked to associate with the robot's behaviours one of Ekman's six basic affective states (joy, sadness, disgust, fear, anger, and surprise) and rate their own confidence in the choice on a 5-point Likert scale [1 = “Not at all" to 5 = “Very much"]. At the end of their interaction, we assessed their overall perception of robot by asking them whether they would like to interact with a physical robot on a scale from 1 [“Not at all"] to 7 [“Very much"].

4.2.2 Results

An a priori sample size calculation using G*Power considering ANOVA as analysis ($\alpha = 0.05$, power = 0.95, number of groups = 3, number of measurements = 5), and moderate effects (f(V) = 0.25), resulted in a sample size of 96 participants. We recruited 102 participants, of which 54 identified as male and 48 as female, aged between 18 and 66 years old ($M = 36.45$, $STD = 13.99$). The majority of participants (79.40%) lacked any experience with robots; 7.80% were programmers or researchers, while the remaining people had primarily experienced robots through television, social media, or exhibitions. Participants' experience with robots included Furhat, Pepper and Roomba robot. Each participant was assigned to one condition, and they were overall distributed among the three experimental

yc72srkp

[4] The animations used in *C3* condition can be viewed on https://tinyurl.com/ycynm64d



(a) Condition 1: only pose.

(b) Condition 2: pose and facial emotions.

(c) Condition 3: pose, facial emotions and sounds.

Figure 4.7. A heatmap for the affective expressions associated with the robot's behaviours by the participants. Colours range from low scores in red to high scores in green.

conditions as follows: 1) 33 people in the only-face (*C1*) condition, 2) 32 participants in the pose-face (*C2*) condition, and 3) 37 participants in the pose-face-sound condition (*C3*) condition.

The results allowed to classify a set of modalities for expressing robot emotions (see Figure 4.7). Note that 4.7a and 4.7b share a common y-label.

The heatmap in Figure 4.7a shows that participants were able to correctly recognise sadness and surprise emotions with the robot's pose showed. They were more undecided in associating the robot's poses to the disgust, joy, and fear emotions, even though we can observe that they were confused with emotions having similar polarity. These results also show that anger was the only emotion completely misinterpreted in the condition *C1*.

In conditions *C2*, participants correctly associated the robot's animations with the disgust, joy, anger, surprise, and sadness emotions (see Figure 4.7b). Observing this heatmap, we can also notice that both animations representing fear were not as clear as the others. Interestingly, previous studies [159, 154] also highlighted the difficulty for participants to recognise it as expressed by most robots.

As expected, the sounds used in condition *C3* allowed participants to almost uniquely associate an emotion to the robot's animation showed (see Figure 4.7c). Finally, at the end of the study, participants were asked to express their desire to use in person the robot. The majority of participants (76.00%) stated that they would like to interact with a physical *ClassMate*, the 9.00% of participants were unsure if they would like to interact the *ClassMate*, and the remaining expressed a negative response.

Regarding **RQ1.2** (How can robots display complex social cues, such as emotions, to enhance the depth of their social presence?), the findings of this study suggest that emotions are an effective way of displaying social cues even in non-humanoid robots. However, if emotions are correctly identified with single modes, little improvement is obtained by the use of additional ones. This result is in contrast with the one from Löffler *et al.*[109] that show how expressing emotions in a multimodal way increases their transparency.

Our result, reflects well the minimalism design principle "less is more" as suggested by Ludwig Mies van der Rohe.

Limitations Despite the valuable insights provided by this study, limitations should be acknowledged to guide future research. The *ClassMate* robot employed in this study was one of the first prototypes of its kind. This has led to develop only a limited range of emotional expressions due to the kinematic and motion constraints. To address this limitation, future work could employ similar robot platforms with enhanced capabilities for expressing emotions. Another limitation can be found in the participants employed. In particular, the convenience sample relied on native Italian speakers that might perceive emotions in a homogeneous yet biased way. Age, gender, cultural background, or other demographic variables might shape the way we perceive the robot's emotional expressions. For this reason, in order to improve the external validity of this result, a possible solution would be to diversity the recruited participants.

Overall, this chapter tackled how robots can display social cues (**RQ1**) using first an AMR using lights and sounds to display simple social cues, second by developing multimodal emotional expressions intended as complex social cues on a non-humanoid robot. These results can be used to build robots for a variety of applications with the focus on how they can display social cues.

Robot applications designed for our social environment require the robot to also conduct social interactions. These can be modelled as spontaneous according to Spontaneous Interaction State Machine (SISM), therefore it is important to study how the robot can use its social cues purposefully. For instance, how can robots purposefully initiate interactions?

Chapter 5

Using Social Cues for Starting Interactions

It is undoubtedly hard to establish the exact moment an interaction starts. That split second that allows an acquaintance to initiate a conversation with us without our full acknowledgement of their presence. Considering Human-Human Interactions (HHIs), interactions can start thanks to a combination of verbal and non-verbal behaviour. Moreover, the environment might nudge a specific communication mode. For instance, in a loud environment, short and effective communication it is likely not done verbally. Given the use-cases that our Society offers for service robots that are capable of interacting with us, this chapter revolves around the following questions: How can a robot exploit its social cues to start interactions? This study specifically addresses how social cues can be used purposefully (**RQ2**) and emphasises their importance in eliciting interactions (**RQ2.1**).

Targeting these questions, this chapter focuses on the following publications:

> Francesco Vigni and Silvia Rossi. Exploring non-verbal strategies for initiating an hri. In International Conference on Social Robotics, pages 280–289. Springer, 2022.
>
> Francesco Vigni, Esteve Valls Mascaro, Dongheui Lee and Silvia Rossi. Between Task and Social Engagement of a Social Service Robot. *Submitted to* IEEE Robotics and Automation Letters, 2024

Vasilis Mizaridis*, Francesco Vigni*, Stratos Arampatzis, and Silvia Rossi. Are emotions important? a study on social distances for path planning based on emotions. In 2024 33rd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). pages 176-181. IEEE, 2024.

5.1 Using Social Space for Starting Interactions

In the context of HHI, the dimensions of social space and the use of non-verbal cues play a pivotal role in conveying intent and fostering engagement among us. The capacity to utilise such cues enables people to traverse social spaces, forge connections, and initiate dialogues without the necessity of explicit verbal interaction. In the context of Human-Robot Interaction (HRI), it is essential for social robots to effectively use non-verbal signals to facilitate interactions that are both natural and intuitive. This section examines the ways in which robots can exploit social spaces to initiate interactions with humans, highlighting non-verbal strategies such as gaze behaviour, approach trajectories. In doing so, this study tackles the question "To what extent, if any, do non-verbal behaviours influence the start of interactions?" (**RQ2.1**).

5.1.1 Methods

In [201], we explore how the approaching policy and gaze behaviours of a humanoid robot can influence the perceived intention to interact before the interaction starts. This scenario mimics well a spontaneous encounter of a robot with a human when physically in the same public space, here represented by a hall (see Figure 5.1).

Participants are instructed to stand still, observe the robot approaching and state the keyword "yes" at the earliest opportunity upon formulating a positive inner response to the question:

(1) "Would the robot like to start an interaction with me?"

A stopwatch is used to measure the duration taken by each participant to say the keyword across the four designed conditions. Two levels of

* co-first authorship

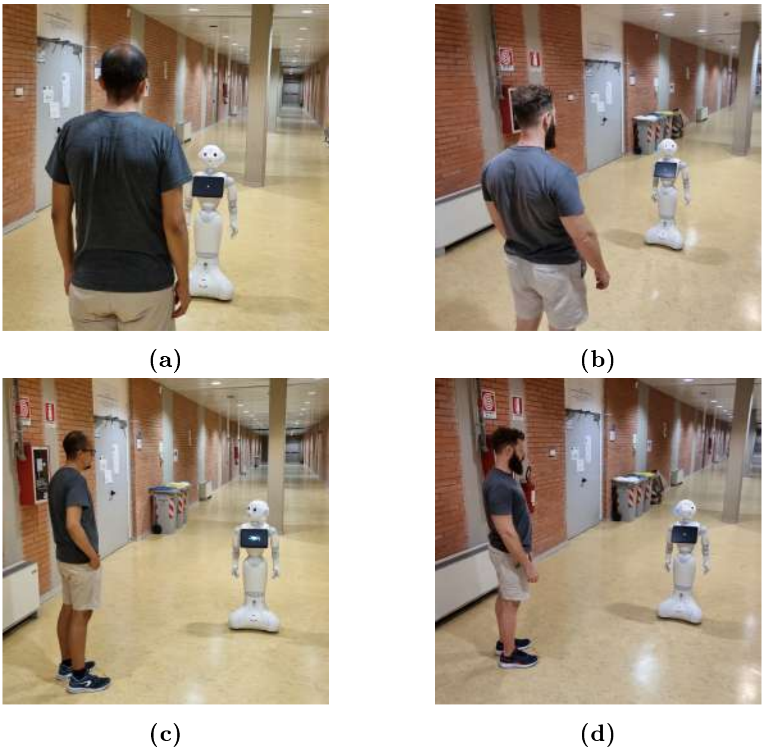

(a) (b) (c) (d)

Figure 5.1. Snapshots of the four experimental conditions.

two variables are considered to build the 2x2 within-subject user study. The scenario allows the robot to undergo three separate phases, namely *Approaching*, *Interacting*, and *Terminating*. During the first phase, the robot is navigating towards the person. The second phase highlights when the robot is interacting verbally with the person, and finally the last phase describes the motion of the robot to return to its original location in the hall. These phases are represented in Figure 5.2 with different colour sections. The experimental conditions are manipulated only during the *Approaching* phase. For the robot's gaze, two conditions are chosen by employing a social or an adverse gaze during the approach phase. In this sense, a social gaze consists in the robot employing a face-directed gaze during the approach phase, while an adverse gaze consists in the robot gazing at the location that is specular to the human's face with respect to the robot's path. Hence, the robot is gazing at an empty location in the hall. The other variable is the robot's approach policy, and is controlled by the position of the standing human in the hall (front vs. side approach). The human can either stand in front of the robot's path or with a lateral offset to it. The obtained experimental conditions are therefore: Social-Front (SF), Asocial-Front (AF), Social-Side (SS), and Asocial-Side (AS). Figure 5.1 shows a snapshot of all the experimental conditions as: Social-Front (SF) is shown in Fig. 5.1a; Asocial-Front (AF) is shown in Fig. 5.1b; Social-Side (SS) is shown in Fig. 5.1c; finally, Asocial-Side (AS) is shown in Fig. 5.1d. After being exposed to each condition, the participants answered a brief post-interaction survey comprising the following 5-point Likert scale entries ranging from 1 (I fully disagree) to 5 (I fully agree). In italics are highlighted the keywords used when reporting the results.

1. The robot's behaviour is *social*.
2. The robot would like to *interact* with me.
3. I would feel *comfortable* of encountering this robot in a social context.
4. I like the *quality* of the robot.
5. I quickly understood when the robot wanted to *start* the interaction.
6. I quickly understood when the robot wanted to *finish* the interaction.

The conducted user study ($N = 26$) reveals that a robot that maintains gaze while approaching enhances the clarity and speed of the perceived intention to interact, compared to direct approaches with adverse gaze.

5.1.2 Results

The responses to question (1) are collected in terms of seconds (time between the start of the robot motion and the keyword pronounced by the participants).

A paired t-test with 95% confidence intervals is performed on the mean of these for each condition. Figure 5.2 shows the mean time to answer the

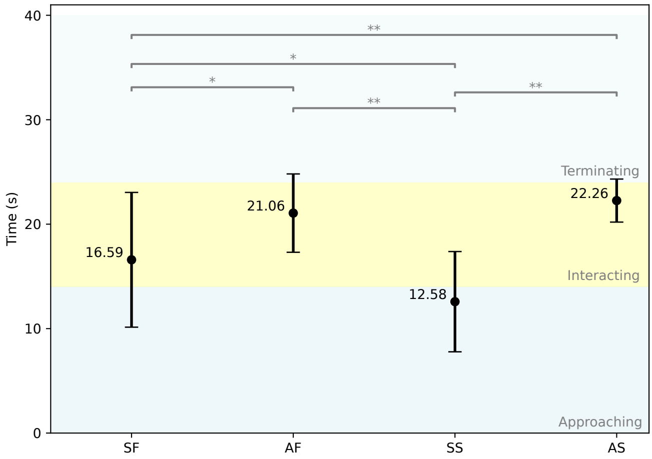


Figure 5.2. Responses' means of question (1) per condition. Significant differences between conditions have been indicated with * for $p < .05$ and with ** for $p < 0.001$.

question (1) and the standard deviation at each condition is shown in terms of error bar length.

In this measure, we found a significant difference between *SF* ($M = 16.59$, *STD* $= 6.45$) and *AF* ($M = 21.06$, *STD* $= 3.75$), with $t(25) = -2.56$, $p < 0.05$; between *SF* and *AS* ($M = 22.26$, *STD* $= 2.06$), $t(25) = -3.76$, $p < 0.01$, and between *SF* and *SS* ($M = 12.58$, *STD* $= 4.80$) with $t(25) = 2.20$, $p < 0.05$. This shows that participants were able to answer significantly faster to the question (1) when the robot employed a social gaze compared to the robot that used an asocial gaze despite its base motion trajectory. A significant difference is found between *AF* and *SS* with $t(25) = 7.84$, $p < 0.01$ and between *SS* and *AS* with $t(25) = 8.27$, $p < 0.01$. Participants took significantly longer to answer question (1) when the robot employed an adverse gaze despite its base motion trajectory. Overall, results show that question (1) was answered significantly faster when the robot employed a social gaze and the user was not in front of the trajectory of the robot.

We could deduce that the base motion trajectory is less relevant than the gaze direction for eliciting the intention to start a social HRI. It is interesting to notice that in Figure 5.2 only the condition *SS* obtained a mean time within the approaching phase window. For completing the study, the post-interaction survey also undergoes the statistical analysis. In particular, a Wilcoxon signed-rank test is performed on the mean responses per each condition regarding the questions in the post-interaction survey. Figure 5.3 shows the mean responses of the survey grouped by conditions with the respective significant differences.

Significant differences are found in the *social* question between SF and AS ($T = 55$, $p < 0.05$), between AF and AS ($T = 48$, $p < 0.05$) and between SS and AS ($T = 18$, $p < 0.05$). Regarding the *interact* question, significant differences are found between SF and AF ($T = 22$, $p < 0.05$) and between SF and AS ($T = 20$, $p < 0.05$). The experiments were designed so that the robot 1) approaches the user in four different ways and 2) terminates the interaction using the same behaviour across all conditions. Surprisingly, no significant difference was found in the *start* question, but a significant difference was found in the *finish* question between SS and AS ($T = 41$, $p < 0.05$). We suspect that this result is given by the short time allocated to the interaction and the difference in the yaw of the robot head between SS and AS.

A significant finding, in the context of the Spontaneous Interaction State Machine (SISM), is the condition SS wherein participants articulated the keyword when the robot had not yet come to a complete stop in front of them but was still in the process of approaching. Hence, the robot was still moving and at around $2m$ distance from the participant.

From [201] we can draw conclusions that are relevant in terms of how spontaneous interactions happen between humans and robots. A first re-

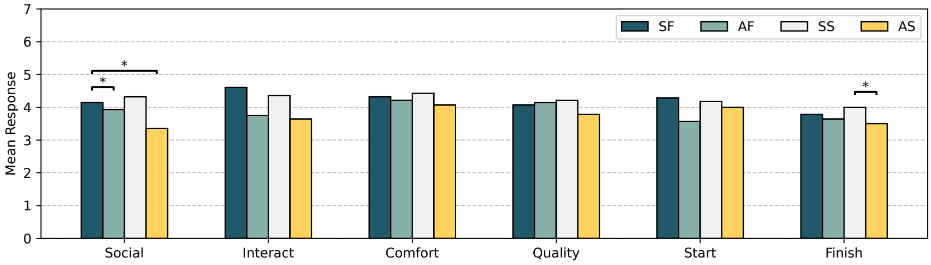


Figure 5.3. Average responses of the survey. Significant differences between conditions have been indicated with * for $p < .05$.

sult in the frame of SISM is that prior to the approach, the robot was located $4m$ away from the participant and no interaction was ongoing. Once approached, the robot engaged participants in a small conversation; hence, an interaction is occurring. Therefore, one may ask, When and how did the social interaction actually start? Which robot behaviours contributed to eliciting it?

Proxemics were crucial to enable social interactions. In this study, despite participants were invited to state the keyword “yes” to the above-mentioned statement, they were not forced to say it at any time. In other words, despite participants answered to that question with different timings, they all answered to it. This can be interpreted by the social valence of the robot interrupting its navigation right in front of them. Regarding which behaviours contributed in eliciting social intention of this robot, our results suggest that gaze behaviour is a stronger social cue with respect to user’s relative spatial position (see Figure 5.2). Yet, gaze behaviour might not be easily readable if the robot is frontally approaching.

Another relevant result lies in the experimental conditions only manipulating non-verbal robot behaviours; hence, these alone were capable of eliciting engagement. The definitions of engagement are still fragmented; however, in the frame of SISM an engagement metric that only considers non-verbal behaviour seems promising for capturing how an interaction initiates. The next section presents a novel engagement model that considers the spatial relationship between social agents.

Some participants declared that in the condition *SS* the robot gives the impression of actively looking for them, whereas in *SF* they were already in the robot’s trajectory and were not able to assess the intentionality of the robot.

Some other participants described the behaviour of the robot during *SS* as secure and more natural, while others suggested that the relative torso pose would have also benefited the study. Additionally, few participants described the behaviour of the robot during *SF* as aggressive and unnatural.

This study introduces an approach to tackle how robot can purposefully use social cues in spontaneous HRIs (see **RQ2**). The focus here is on how robots can use the social space to initiate an interaction.

This topic holds significance considering the various interpretations we can derive from a robot that navigates in our environments [202, 4]. This study specifically addresses the question of the extent to which non-verbal behaviours may elicit interactions (**RQ2.1**). The findings indicate that, even at a considerable distance from the human, the robot's gazing behaviour is interpreted as a social signal, capable of conveying the intention to initiate an interaction. Therefore, the relative spacial configuration between a robot and a person in its proximity matters when considering a potential interaction.

This study aligns effectively with the rationale introduced in SISM as these social cues can be used for starting spontaneous interactions.

Limitations The sample size of the study ($N = 26$) could limit the generalisability of the findings, thereby compromising the extension of conclusions to a wider population. The emphasis on non-verbal cues such as gaze and spatial approach, although informative, disregards the potential effects of verbal interaction, which could influence the interpretation of social intention. For instance, the robot could start speaking to a participant while approaching them.

The experimental setup exclusively evaluated a robot within a hall scenario, which may not adequately reflect the intricacies of more dynamic and densely populated environments. Replicating this study *in-the-wild* could strengthen the results. Finally, given the impact of gaze and approach behaviour, it would be interesting how other robots are perceived in similar scenarios.

5.2 Using Non-Verbal cues for Starting Interactions

In the previous study, a humanoid robot is approaching a standing person in a hall. By varying its non-verbal cues, we studied to what extent participants perceived its social intentions. A possible limitation of the previous work, however, lies in the unaltered context during the interaction. The question that rises is:

- What occurs if the context shifts while an interaction has the potential to initiate?

Considering this question, alongside the research question "To what extent, if any, do non-verbal behaviours influence the start of interactions?" (**RQ2.1**), allows for envisioning a future scenario in which a robot is exposed to a context shift and have to initiate interactions purposefully.

Social robots are already being deployed in several social environments, and the way these robots can tackle these challenges is a hot topic within the research community. These autonomous robots will allocate portions of their time to engage in social interactions, as well as participate in autonomous activities. Therefore, this section describes a study in which the context shifts while the interaction is occurring.

The developed scenario consists of an autonomous social robot that transitions between performing a task autonomously and handing over an object to a person. In the context of SISM, this can be seen as the robot in the *Pre-Interaction* state while performing a task autonomously and transitioning to the *Social Interaction* upon the person arrival.

Following the rationale of SISM, there will be moments in which the robot will be engaged in a social interaction and others in which it will not. These latter moments can be used for employing the robot in autonomous and specific task. Nonetheless, if the robot is performing its task autonomously in an environment that could potentially also be used for social interactions, its social awareness shall be kept [38]. As a result, the robot should be ready to transition to the SISM "Social Interaction" state at any time.

5.2.1 Methods

We designed a study in which the robot acts as a bartender and is initially by itself performing a traditional bartending task. Shortly after starting the task, a person approaches the table as a bar customer would, triggering a context switch. The person is priorly informed 1) that the robot is currently preparing a drink for a previous customer and 2) it has to hand them over an object (a yellow bottle) located on a side tabletop.

A 2x2 within-subjects study was developed, manipulating two variables: "Gaze" and "Task", each with two levels. Specifically, the robot could exhibit either "Social" or "Asocial" gaze behaviour.

A "Social" gaze behaviour is considered when the robot gazes at the person in front of it, e.g., to acknowledge their presence [57], without staring at them [186]. An "Asocial" gaze behaviour is considered when the robot gaze is at the objects that are manipulated, e.g., the bottle, and never at the person standing in front of the robot.

In accordance with the experimental setup, the robot was performing an autonomous task—pouring water into a cup—prior to the participant's arrival. The levels of the "Task" variable were defined as either (1) finishing the task and then handing over the yellow bottle, or (2) stopping the task as soon as possible, handing over the bottle, and finishing the task later.

The obtained conditions are:

- **Social Interrupt.** The robot would employ a socially acceptable gaze while interrupting its task to hand over the yellow bottle.
- **Social Continue.** The robot would employ a socially acceptable gaze but would first finish pouring water in the cup prior to handing over the yellow bottle.
- **Asocial Interrupt.** The robot would employ a socially unacceptable gaze behaviour and interrupt its task to hand over the yellow bottle.
- **Asocial Continue.** The robot would employ a socially unacceptable gaze behaviour and would first finish pouring water in the cup prior to handing over the yellow bottle.

The four conditions are presented in random sequences to each participant. While observing the robot behaviour, participants are asked to consider the following question:

> (2) "How sure are you that the robot is about to hand you the yellow bottle?"

The interaction interface consists of a slider on a web interface, recording the values of the slider cursor paired with the timestamp. The slider units span from 0 (not sure at all) to 100 (very sure). The slider is included only to improve participants' attention. Each participant was communicated that the robot would hand over to them a yellow bottle. With this instruction, we instil expectations into participants [101].

The following hypothesis were defined:

- **H1** "A robot employing a social gaze behaviour leads to higher sociability and animacy with respect to using asocial gaze behaviour"
- **H2** "A robot initially ignoring user's expectations leads to higher disturbance with respect to satisfy them right away"
- **H3** "A robot complying to user's expectations leads to higher sociability with respect to postponing them"

After observing the robot behaviour, participants were asked to respond to a survey comprising the factors of Human–Robot Interaction Evaluation Scale (HRIES) [183] (Sociability, Animacy, Disturbance, and Agency).

We plan to analyse the results using ANOVA for this within-subjects study. Therefore, the desired sample size is $N = 50$ and is computed with the desired power of 0.80, the significance level $\alpha = 0.05$, and the effect size of 0.25, considering as statistical test as "ANOVA: Repeated Measures, within factors"[1].

A pilot study is conducted with $N = 20$ participants that observed the robot in real life in the premises of the Technical University of Vienna, Austria. The methods used to run this pilot study are described in Subsection The Real World and serves to improve the robot's behaviours. The user study is conducted on $N = 50$ participants that observed the robot in a virtual environment described in Subsection The Virtual World. The interaction scenario for the real world study is visible in Figure 5.4, while a screenshot of the study conducted online is visible in Figure 5.6

The Real World

The dual-arm version of the Tiago robot by Pal Robotics[2] was located behind an L-shaped table, resembling the tabletop of a common bar (see Figure 5.4). The robot is instructed to pour water into a cup using its right arm and to hand over a yellow bottle with its left arm to an upcoming person. The robot interacted with the environment in a pre-scripted and

[1] G*Power 3.1.9.7

[2] https://pal-robotics.com/robot/tiago/

non-reactive way. We used the motion planning framework MoveIt[3] and exploited its built-in strategies for the manipulation tasks.

The tables' positions were manually added to the planning scene[4], and the poses of the objects to be manipulated were carefully obtained empirically.

This strategy allowed for a fast and effective implementation of pick-and-place using inverse kinematics while ensuring collision free arms trajectories. Moreover, given the cylinder-like shapes of all the objects to be grasped, the Pal Robotics Parallel Grippers with only 2 Degrees of Freedom (DoF) were sufficient to perform the manipulation tasks.

A touchscreen tablet is located on top of the table at an ergonomic position so that it can be easily reached by participants. The tablet implemented a User Interface (UI) displaying a slider with values between 0 and 100 (see Figure 5.5). Participants were instructed to control the cursor of the slider whilst observing the robot's behaviours. Moreover, participants received a brief overview to the scenario, during which they were informed that: "The robot is instructed to hand you a yellow bottle once you stand in front of it." The buttons (Start, Stop, and Reset) on the UI were simply for debugging purposes and participants were instructed not to interact with them.

[3] https://moveit.ai/

[4] planning_scene

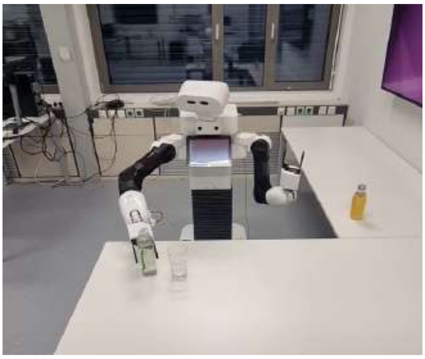
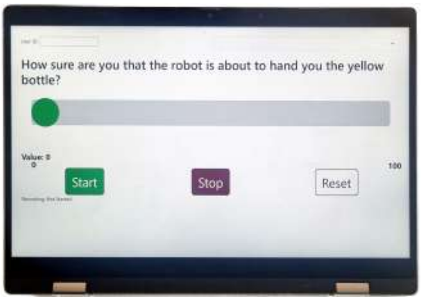


Figure 5.4. Bartending scenario with L-shaped table and dual-arm Tiago robot.



How sure are you that the robot is about to hand you the yellow bottle?

Figure 5.5. User Interface (UI) for capturing the social intentions of robot bartender.

The implementation of the logic to record the data was such that: 1) a timer started in sync with the robot's behaviour, 2) every time the cursor was moved, a new tuple (timestamp, slider-value) was added to a vector; and 3) once the video terminated, a `.csv` file was stored in disc. Feedbacks from participants were collected and resulted in rephrasing the question presented with the slider from “How sure are you that the robot is about to hand you the yellow bottle” to a more concise “The robot is performing the handover of the yellow bottle” (see Figure 5.6).

The Virtual World

Four videos of 50 seconds each are recorded and shown to the participants on a custom-web-based platform for running the experiment. The platform is powered by Flask[5] and is hosted using the free tier option of the service PythonAnywhere[6]. The platform is designed to allow running the study only on devices with large screens (at least 1200px of width). Moreover, cookies are stored in the browser to enforce the within-subject design, in which each participant is exposed to all conditions once. The platform sequentially shows the videos to the participants, displays a slider that is controllable with the cursor of the participant's device, and proceeds in administering the questionnaire. The platform shows the various experimental conditions in a pseudo-random fashion and selects the sequence of conditions for upcoming participants, selecting the least frequent sequence of conditions already presented to other participants. In this way, it is possible to run the study online and obtain a balanced distribution of condition sequences.

A logic written in JavaScript allows the platform to play the videos while concurrently starting a timer to track how the slider is changed along the time. A custom relational database is designed and implemented with SQLite[7]. This database manages the experiments, stores the time-value of the sliders and is accessible only via a private key.

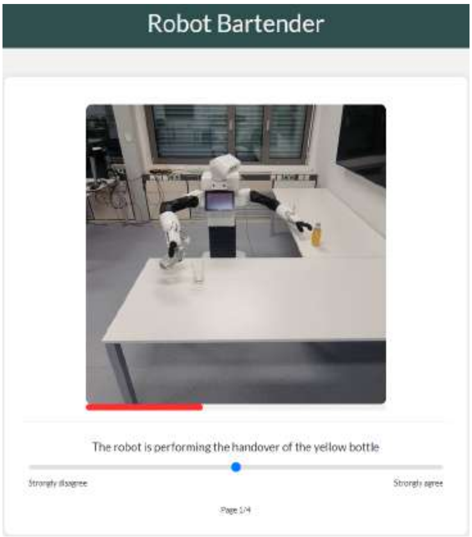
The overall implementation is publicly available at [8].

[5] https://flask.palletsprojects.com/

[6] https://www.pythonanywhere.com/

[7] https://www.sqlite.org/

[8] github.com/vignif/virtual_bartender



Robot Bartender

The robot is performing the handover of the yellow bottle

Strongly disagree

Strongly agree

Page 1/4

Figure 5.6. Screenshot of the virtual interface displaying a video and slider for assessing robot's behaviour (see The Virtual World).

5.2.2 Results

To analyse separately the manipulated variables, we grouped the results in terms of tasks and gaze. Figure 5.7 shows the mean responses to HRIES grouped per gaze, while Figure 5.8 shows the mean responses per task variations. Considering **H1**, we grouped the answers to HRIES according to the variable of interest, here is the gaze behaviour. The paired t-test does not indicate statistically significant differences with respect to the sociability ($t = 1.42$, $p = 0.16$) and animacy ($t = -0.44$, $p = 0.67$) factors (see Figure 5.7). Therefore, **H1** is rejected.

Considering **H2**, we grouped the answers to HRIES according to the variable of interest, here is the task policy. The paired t-test does not show any significant differences in the disturbance between the conditions in which the robot continued its task, with respect to when it interrupted it to hand over the bottle to the participant (see Figure 5.8). Suggesting that participants were overall pleased by the robot across all the scenarios, hence, rejecting **H2**.

Finally, considering **H3**, we tested the perceived sociability of the robot

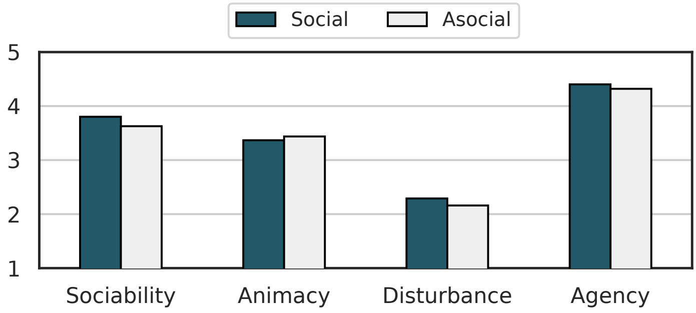


Figure 5.7. Responses to post-interaction survey Human–Robot Interaction Evaluation Scale (HRIES) grouped per gaze conditions.

along the different types of task policies. In particular, as similarly done for testing **H2**, we grouped the answers to HRIES according to the task policy. The paired t-test shows a significant difference ($t = -2.48$, $p = 0.01$) in the sociability factor between the variations of tasks. The "Interrupt" task policy of the robot was able to convey higher sociability with respect to the "Continue" task policy. This result allows accepting **H3**.

This latter result was designed as a confirmatory hypothesis that expects participants to perceive lower sociability when the robot ignored their expectations (Task policy "Continue") with respect to when the robot interrupted its task (Task policy "Interrupt") in order to perform the hand-over of the yellow bottle.

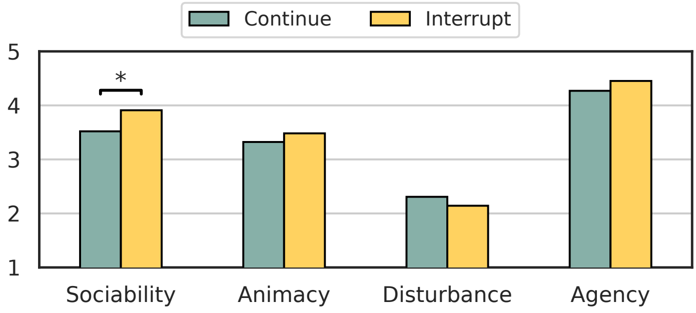


Figure 5.8. Responses to post-interaction survey Human–Robot Interaction Evaluation Scale (HRIES) grouped per task conditions.

On the Attention Span Participants were instructed to control the cursor of a slider in a UI (see Figure 5.6) while being exposed to the experimental conditions. The slider is designed to implicitly measure participants' understanding of the robot task. Preprocessing is needed since the slider data are as frequent as the participant input on the cursor. The preprocessing consists of dividing in bins each trial and average the values of the slider within each bin. In this way, the records from different participant would match in sample size. We set the bin size in the most conservative way, so by setting it to the highest frequency in the raw data per each condition. A small bin size result in averaging a lower number of samples per time. Therefore, this strategy attempts to preserve the trends in the data, yet, it also preserves the noise in it.

Figure 5.9 shows the data from the sliders across the four experimental conditions. Notice that all the subfigures start with a value of 50 units due to the initialization of the slider in the UI. It can be appreciated that 1) the conditions for which the robot interrupted the task ("Social Interrupt", and "Asocial Interrupt") shows very similar trends. Similar rational can be seen between the conditions for which the robot continued the task.

To understand how similar the data collected with the sliders are along conditions that employ the same task policy, we computed their Pearson's correlation coefficient. Figure 5.10 shows the correlation matrix obtained with the defined coefficient. It is immediate to grasp the similarity between the conditions that employ the "Interrupt" task policy. A similar result can be seen when considering the conditions that employ the "Continue" task policy.

The study explored different contexts, testing whether a robot's social or asocial gaze, combined with the performed task, influences participants' perceptions about the robot's actions.

Results show that although the robot's gaze behaviour was noticed by participants, the impact on their perception of the robot's sociability and animacy was not statistically significant. Assessing **RQ2.1** (To what extent, if any, do non-verbal behaviours influence the start of interactions?) we can conclude that the autonomous robot task was correctly identified by participants as shown in Figure 5.9. From this figure, we can notice two effects. First, the levels of the task variable are correctly identified and show a horizontal shift between conditions in which the robot interrupted

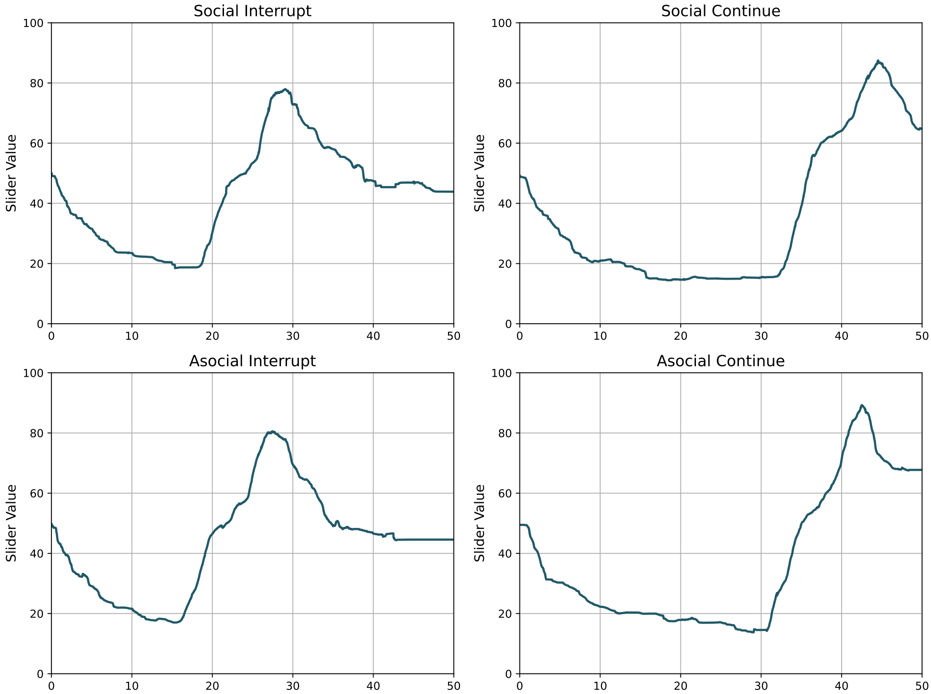


Figure 5.9. Plots of the slider values resulted from the virtual scenario grouped per each experimental condition.

its task to when it continued it. Second, the task assessment performed with the slider "the robot is performing the handover of the yellow bottle" (see Figure 5.6) is not affected by the variations in the gaze variable.

This suggests that while non-verbal behaviours are crucial, other factors may moderate their effectiveness in eliciting meaningful HRIs. Starting an interaction can rely on both verbal and non-verbal cues, however, if the robot is simply a mobile platform, the interaction interfaces are limited and interaction might simply mean to successfully convey navigational intent.

Limitations Despite the strengths of this study, several limitations should be acknowledged. One limitation is the use of a fixed task scenario. The study consisted of a predetermined and structured task for the

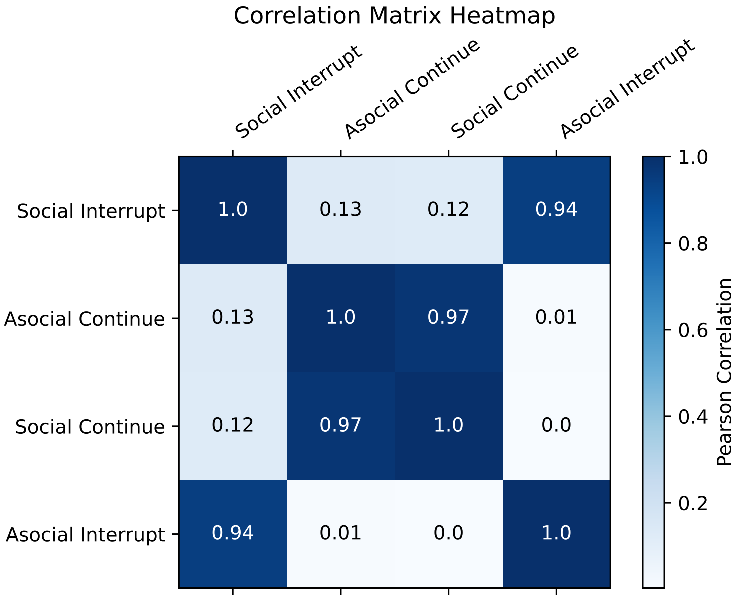


Figure 5.10. Correlation Matrix Heatmap of the slider values grouped per experimental condition.

robot, which constrains the applicability of the results to more fluid and unpredictable real-world contexts. The robot's behaviour was limited to particular actions, such as pouring water or passing an object, which may not adequately reflect the adaptability needed for more intricate interactions. Another limitation lies in the focus on non-verbal communication. While the study provides valuable insights into non-verbal cues in HRI, it does not account for the potential role of verbal communication. In real-world interactions, people often rely on a combination of verbal and non-verbal signals to coordinate tasks and engage socially. The absence of verbal communication in the study narrows the scope of the findings, making it difficult to generalize the results to more realistic interaction scenarios.

Finally, the study involved a certain level of deception, as participants were not fully informed about the robot's capabilities. This methodological choice, while necessary to maintain experimental control, may have influenced participants' perceptions and behaviour. Deception could affect how participants engage with the robot, limiting the ecological validity of the study's findings. This factor should be considered when interpreting the results, particularly in terms of how participants perceive and interact with robots in less controlled, real-world settings.

5.3 Using Emotions for Adapting Social Cues

Thus far we have seen how social robots with various designs can control their behaviours so to start interactions. The range of possible spontaneous interactions however is far from being fully addressed here. Human interactions are frequently influenced by emotional states, ranging from subtle gestures of empathy to uncontrolled displays of joy or anger. Context, emotional states and other factors can influence how we perceive and use the surrounding space [44]. Therefore, our perception of the social context is significantly impacted by the emotional states of those around us [118]. Moreover, the appropriateness of different robot navigation behaviours (approaching, not moving or moving away) has been shown as linked to the observer's emotional states [147].

For this reason, we conducted a study using a popular Autonomous Mobile Robot (AMR) like the iRobot Roomba to investigate whether emotions elicited from a loudspeaker, affect human perception of robot proximity. In doing so, we frame this scenario as the robot starts in the *Pre-Interaction* and approaches to the *Social Interaction* state of SISM. To contribute in this direction, this section ponders whether *robots should adapt their behaviour according to the emotional states of the people they encounter*. In particular, again tackling the research question "To what extent, if any, do non-verbal behaviours influence the start of interactions?" (**RQ2.1**), this study explores the role of human emotional states given a robot approaching them. While some argue for a uniform approach where consistency promotes predictability and trust, others claim that the richness of human emotion requires a more nuanced response and robots are not yet capable of reproducing it or adapting to it effectively [47, 35].

Each participant observed the robot approaching them in two different encounters. The first time, we asked the participants to use a remote controller to stop the robot at a preferred minimum distance (pmd), while the second time, the robot computed a path to avoid them according to

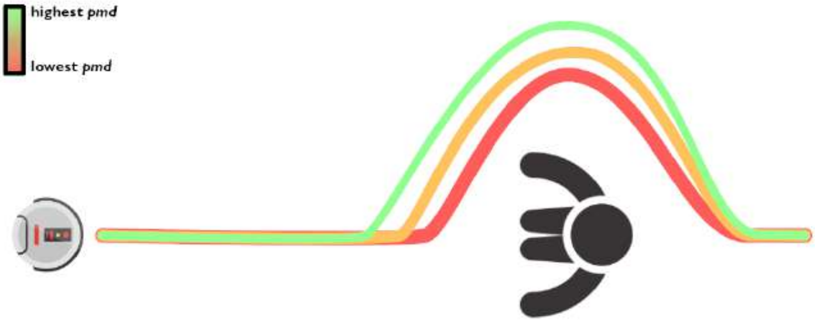


Figure 5.11. Top-view sketch of the second encounter with the Autonomous Mobile Robot (AMR) in which the paths are computed considering the preferred minimum distance (pmd) obtained during the first encounter.

the preference set in the first encounter. Emotions are induced by using sounds from an external speaker in the hall.

5.3.1 Methods

The scenario started with a participant standing about $3m$ from the popular AMR Turtlebot4 lite[9] in a hall. The participant was handed a remote controller and instructed to observe the robot approaching them.

Figure 5.11 sketches the top-view of the second encounter. The colours in the figure are designed to closely follow the logic on the pmd, as shown by the legend in the figure. Notice that the robot paths represented in Figure 5.11 simply sketches three exemplary paths, obtained by picking three Pmds.

With the remote controller, participants were able to initiate the robot approach and halt its motion according to the individual's preference (pmd). Upon starting the experimental scenario, the external speaker played a sound utterance according to the experimental condition. Then, the participants had to initiate the robot's approaching motion by pressing the button Y on the remote controller. This approach is inspired by Cook *et al.*[43], who demonstrated that different music genres play differ-

[9] https://clearpathrobotics.com/turtlebot-4/

ent roles in emotion regulation, contributing to either positive or negative emotional experiences. Therefore, we elicited an intended emotion (positive or negative) in participants according to the sound utterance. This choice makes it possible to simplify the emotion recognition problem to its extreme valence values.

The two experimental conditions are defined as:

- The participant is elicited with a **Positive emotion** via the sound utterance;

- The participant is elicited with a **Negative emotion** via the sound utterance.

The positive sound utterance was a recording of a baby babbling and laughing, while the negative emotion was a recording of a baby crying, publicly available[10]. This choice is made starting from the neuroscience literature on aversive and inviting stimuli and their link to emotional regulators [145].

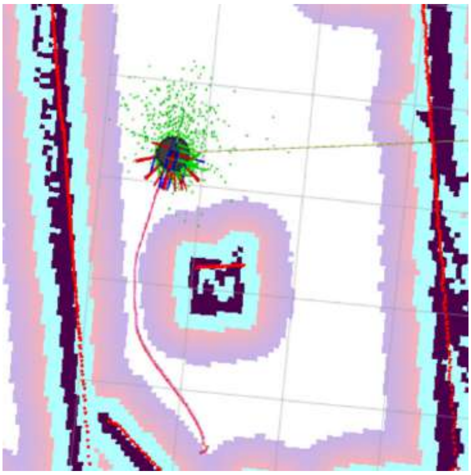
The maximum speed of the robot was set to $0.20m/s$. By pressing X on the remote controller participants were free to stop the robot's motion, implicitly setting their pmd to it. Its low speed ensured low error in the measured distance when commanded to stop. At this point, the approach terminated, the speaker silenced, and the robot initiated a path to its initial pose. If the participant decided never to press the stop button on the remote control, the robot was instructed to stop at a minimum fixed distance of $0.1m$.

At this point, the robot was informed of the participant's pmd and was instructed to navigate from the starting position at $3m$ to a point at $1m$ behind the participant, avoiding the collision.

Again, the robot waited for the participant to press the start button Y, to start the navigation. Throughout this interaction, the external speaker was off. The robot proceeded in computing and performing a path, which was modulated to take the pmd into account for the parts of the path that were closer to the participant (see Figure 5.12). For example, if a participant set that the pmd to $0.5m$, the robot would have computed a path while avoiding the participant with a minimum distance equal to

[10] https://tinyurl.com/4jszu78a

Figure 5.12. Example of robot's path during the second approach.



$0.5m$ (pmd). This functionality was achieved by modifying the inflation layer of the navigation stack used by the robot to compute the path to the given goal autonomously.

The Ethics Committee of the University of Naples Federico II approved the within-subjects user study reported here. The study was conducted at Noosware VB[11] premises in Eindhoven, The Netherlands. All participants were exposed to both experimental conditions only one time. To reduce order effects, the order in which participants were exposed to the experimental conditions was counterbalanced.

An a priori power analysis was conducted, and to achieve an effect size of 0.50 with a statistical power of 0.88 and a significance level of 0.05 a sample size $N = 34$ was required. This analysis was performed by selecting as a statistical test the "difference between two dependent means"[12], as the same participant is exposed to both experimental conditions.

Consistent with the Research Question (RQ) introduced above, this user study focuses on the following hypotheses:

- **H1**: Emotions, whether positive or negative, might result in a different pmd to stop the approaching robot (exploratory hypothesis);

[11] https://noosware.com/

[12] G*Power 3.1.9.7

- **H2**: Participants might have a different perception of the robot's avoiding trajectory, considering a positive or a negative emotional reaction (exploratory hypothesis).

We included all adults and healthy participants with no declared impaired hearing, while participants with declared healthy-related issues were excluded from the study. This approach was adopted to ensure that the experiment would yield reliable and unbiased results.

Our convenience sample was drawn from university staff and students. A total of 34 participants were recruited, of whom 8 self-identified as female and 26 as male. Age ranged from 22 to 57 years ($M = 27.50$, $STD = 5.84$).

Participants were asked to assess their experience with robots on a scale from 1 (no experience at all) to 7 (very experienced).

After 1) observing and implicitly setting on the pmd, and 2) observing, for the second time, the robot navigating to a point behind them, participants were asked to fill out a survey for calculating the HRIES [183]. The survey was augmented with questions about prior experience with robots, demographics, and the following entry: "Considering the last robot path, how closely are these sentences with you? 1 (not at all) - 7 (totally)":

1. The robot maintained an appropriate distance
2. The robot moved too close to me
3. The robot moved too far from me

Considering these questions, we expect to observe an inverse relationship between (2) and (3) as suggested by their semantics.

The robot used was shipped *Humble* Robot Operating System (ROS) version[13] (ROS2). A laptop with Ubuntu Jammy Jellyfish (22.04) was orchestrating the communication using the same robot ROS version. Dedicated software was written to:

- Read the inputs from the remote controller *Speedlink rait*[14] (connected via USB-A)
- Control the navigation of the robot as per the interaction scenario

[13] https://docs.ros.org/en/humble/index.html

[14] https://www.speedlink.com/

Table 5.1. Mean and standard deviation of the preferred minimum distance (pmd) grouped per condition.

	Mean	STD
Positive	0.87	0.31
Negative	0.90	0.30

- Store in a `.csv` file the pmd as set by the participant during the first robot approach
- Control the sound played by the external speaker (connected via Bluetooth)

The distance stored in the `.csv` file was measured in meters and acquired via the onboard robot Light Detection and Ranging (LiDAR) sensor during the first robot approach. For computing the path needed for the second approach, this file was read, and the information was fed to the navigation stack Nav2[15] to adjust the robot path coherently.

5.3.2 Results

Based on previous research on HHI, we analysed participants' responses to robots displaying different behaviours. Participants observed the robot's approach while experiencing positive or negative emotions. Our findings suggest that emotional states induced by external stimuli can affect participants' perception of robot proximity. In detail, the results indicate that while comfortable stopping distances were unaffected by participants' emotional state, individuals who experienced positive emotions judged the same proxemics distance used while performing an avoidance behaviour to be more acceptable compared to the case of negative emotions. This study describes the extent to which our emotions can alter the perception of robot behaviours, ultimately affecting our acceptance of these novel social agents while they navigate among us. To analyse our results, we first performed a paired t-test on the minimum interpersonal distances collected (means and standard deviations are reported in Table 5.1).

[15] https://navigation.ros.org/

The statistical test shows that the controlled variable had no significant effect on the minimum interpersonal distance, as expressed by the participants' input during the robot's first approach.

To assess the reliability of the factors of HRIES [183] we performed their Cronbach's alpha and show the results in Table 5.2. The factors "Sociability", "Disturbance" and "Agency" are considered reliable as Cronbach's alpha of their sub-items is greater than 0.70, which is the commonly accepted value in the community [82]. Furthermore, attempting to increase Cronbach's alpha by only considering a subset of the "Animacy" items did not yield successful results. No further results are reported for the HRIES factor "Animacy" as it cannot be considered a reliable composite score.

To further the investigation, we examined whether the responses to the HRIES [183], whose factors were found to be reliable, differed between the two experimental conditions. We tested the assumptions for performing the paired samples t-test and concluded that it can be performed on the "Sociability" and "Agency" factors, while the Wilcoxon signed-rank test is the appropriate one for the "Disturbance" factor.

Figure 5.13 shows the means of the responses to HRIES grouped by condition, but no statistical differences are found between the conditions. Regarding the three questions related to the appropriateness of the robot motion, the assumptions to perform the paired sample t-test are not met, so we resort to the Wilcoxon signed-rank test. The average of the responses is shown in Figure 5.14. Interestingly, participants rated with high scores the statement: "the robot maintained an appropriate distance", and a significant difference between its mean values per condition is obtained ($Z = 107$, $p < 0.01$) with a higher evaluation of the appropriateness of the robot's avoiding behaviour in the case of a positive emotional state. The data indicates that participants perceived the distance of the robot to

Table 5.2. Cronbach's alpha of the factors of Human–Robot Interaction Evaluation Scale (HRIES) per condition.

	Sociability	**Disturbance**	**Agency**	**Animacy**
Negative	0.78	0.82	0.81	0.52
Positive	0.74	0.77	0.83	0.55

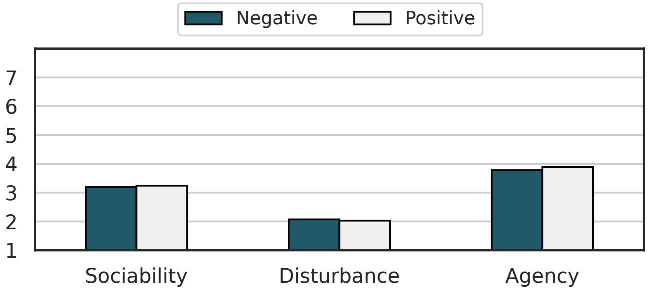


Figure 5.13. Mean responses to the Human–Robot Interaction Evaluation Scale (HRIES)'s factors grouped per condition.

be more appropriate when they experienced a positive emotion compared to when they experienced a negative one. This result gains interest, considered alongside the result from Table 5.1 that reports the average pmd as expressed by participants via the remote controller. On one hand, the data collected by the robot does not exhibit any statistical difference between the condition (see Table 5.1), on the other hand participants rated significantly different the "appropriateness" of the behaviour of the robot during the second encounter as shown in 5.14.

This study reveals insights into the influence of emotional states on HRI dynamics, particularly concerning proxemics. In a previous study by Spatola *et al.*[183], Cronbach's alpha values for the factors of the HRIES were found to be 0.93, 0.88, 0.81, and 0.74 for Sociability, Disturbance, Agency,

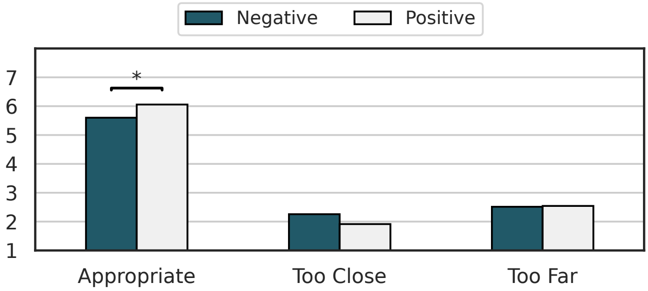


Figure 5.14. Mean responses to the survey questions grouped per condition, significant differences have been indicated with * for $p < 0.05$.

and Animacy, respectively. In our analysis, as seen in Table 5.2, we found that the results for Sociability, Disturbance, and Agency were consistent with those of the previous study. However, the results for Animacy did not yield a reliable composite score. Animacy encompasses traits such as *human-like*, *real*, *alive*, and *natural*. Further analysis of the participants' responses revealed that *human-like* and *alive* traits negatively affected the Animacy score. The study's findings suggest that the Turtlebot4-lite's lack of human-like features and unchanging behaviour may have affected how alive participants perceived it to be, highlighting the need to improve how we assess Animacy. Participants with prior experience with robots (greater than 4 - from 1 to 7) kept on average a greater distance ($1.22m$) from the robot, compared to less experienced participants ($1.19m$), similar to the work of Takayama *et al.*[189]. Despite the heterogeneous self-assessed prior experience with robots ($M = 4.51$, $STD = 1.85$), the feedback gathered from the post-interaction survey revealed that positive emotions made the robot's path more acceptable, even though the trajectory was similar between the two conditions. This finding suggests that emotional stimuli can influence individuals' subjective assessment of robot behaviour (sustaining hypothesis $H2$) when linked to proximity but not to social distances *per se* (not sustaining hypothesis $H1$). Indeed, contrary to our initial hypothesis $H1$, the statistical analysis of minimum interpersonal distances set by participants during the first robot approach did not reveal significant differences between positive and negative emotions. This discrepancy between participants' objective behaviour and subjective perceptions is consistent with De Houwer's research on implicit measures [48].

In this study, the AMR approaches a person standing in a hall, transitioning from the *Pre-Interaction* to the *Social Interaction* state according to SISM. Moreover, this section has described a study that tackles the question "To what extent, if any, do non-verbal behaviours influence the start of interactions?" (**RQ2.1**). Results show that non-verbal behaviours such as adopting navigational paths that are modulated by users emotions can benefit the perceived appropriateness of the robot (see Figure 5.14).

In particular, it shows that people with different emotional states (positive to negative valence according to [165]) perceive an AMR that approaches them differently in terms of the appropriateness of its path. Similarly, other works have investigated the complexity of emotions in HRI and found differences in implicit and explicit measures [191], [187]. This study provides a solid example of how emotions can be used implicitly by an AMR approaching a standing person in the attempt to instrument with some sort of empathy social robots [117].

Limitations While the findings of this study offer valuable insights into the influence of emotional states on proxemics in HRI, limitations should be acknowledged. First, the robot used in this study, a Turtlebot4-lite, has limited social expressiveness due to its lack of anthropomorphic features. As a result, participants may have perceived the robot more as a machine than as a social agent, in line with the reliability of the "Animacy" factor in the HRIES (see Figure 5.2). Future studies could use more socially expressive robots to better improve the animacy of it.

Additionally, the study only considered the valence (positive vs. negative) of emotional states and did not explore the potential role of emotional arousal (intensity). High-arousal positive emotions (e.g., excitement) might affect participants' reactions differently than low-arousal positive emotions (e.g., calmness), as has been suggested in previous research on emotions and social behaviour.

Finally, the participant sample size, though sufficient for detecting some effects, limits the generalizability of the findings. A consideration of larger and more diverse populations would facilitate an in-depth investigation of individual differences, including personality traits and cultural backgrounds, which may influence the relationship between emotions and proxemics in HRI.

Overall, this chapter has tackled the intricate dynamics of how robots can effectively utilize social cues, particularly focusing on non-verbal behaviours, to initiate spontaneous interactions. This exploration is grounded in our overarching **RQ2**: How can robots purposefully use social cues in spontaneous HRI? The studies are primarily centred on the RQ "To what extent, if any, do non-verbal behaviours influence the start of interactions?" (**RQ2.1**). We identified key factors that can significantly impact the way interactions are triggered, such as gaze, context, and users' emotional state. The results highlighted the importance of specific cues in drawing users into interaction, thereby affirming the hypothesis that non-verbal communication serves as a crucial mechanism for spontaneous interactions. This nuanced understanding not only contributes to the existing body of knowledge in HRI but also paves the way for future research on refining robotic communication strategies. In conclusion, the findings presented in this chapter highlight the capacity of robots to operate as socially aware agents, adept at responding to human signals in significant manners. As we advance, the implications of these findings will guide the development of robots that are not only operational but also socially proficient, enabling more harmonious and effective interactions between humans and robots.

Chapter 6

Using Social Cues for Maintaining Interactions

Considering goal-oriented interactions as defined by Perceptual Control Theory (PCT) [120], it is possible to model possible interaction goals for the robot. For example, the robot may be programmed to convey information verbally to the user. Upon the completion of this task, the goal of the interaction will be considered accomplished. Alternatively, the robot may be programmed to maintain the interaction within defined limits. In this context, it is essential for the robot to be equipped with the ability to maintain interactions through its behaviours.

This chapter builds upon the previous one regarding purposeful use of social cues by robots in spontaneous Human-Robot Interactions (HRIs) (**RQ2**) in two distinct user studies. The first exploit a type of interaction that is natural for humans: verbal communication.

The field of Natural Language Processing (NLP) has informed machines like robots about the rules of our natural languages. More recently, Large Language Models (LLMs) and off-the-shelf tools have promoted the integration of conversational agents into social robots [42]. This is obtained by also addressing ethical concerns related to non-verbal cues, misinformation, emotional disruption, and biases [119]. With this in mind, the research question "To what extent, if any, do different robot's communication styles maintain interactions?" (**RQ2.2**) is addressed in a user study conducted *in-the-wild* in which a robot engaged users in a quiz game using different communication styles. The second study is presented in this chapter and tackles the subquestion "To what extent, if any, do different robot's emotional-adaptive behaviours maintain interactions?" (**RQ2.3**).

The study involves a robot adjusting the interpersonal distance based on the user's emotion while holding a free conversation with them. In this latter study, the robot is instrumented with a primitive form Emotional Intelligence (EI) that allows it to change its behaviour (manipulate interpersonal space) according to the user emotional state. As such, it contributes to addressing the challenges on emotional intelligence in robotics as presented in Marcos-Pablos *et al.*[117] about robots' ability to express empathy.

This chapter encompasses the following publications:

> Francesco Vigni, Antonio Andriella, and Silvia Rossi. Sweet Robot O'Mine - how a cheerful robot boosts users' performance in a game scenario. In 2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), pages 1368–1374. IEEE, 2023.

> Francesco Vigni, Dimitri Maglietta, and Silvia Rossi. Too close to you? a study on emotion-adapted proxemics behaviours. In 2024 33rd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). pages 182-188. IEEE, 2024.

6.1 Using Communication Styles for Maintaining Interactions

The ability to impact the attitudes and behaviours of others is a key aspect of Human-Human Interaction (HHI). The same capability is a desideratum in HRI, when it can have an impact on healthy behaviours. The robot's interaction style plays a significant role in achieving effective communication, leading to better outcomes, improved user experience, and overall enhanced robot performance. Nonetheless, little is known about how different robots' communication styles impact users' performance and decision-making. This section describes and elaborates the results of [196] in which the main focus is on **RQ2.2** that states: "To what extent, if any, do different robot's communication styles maintain interactions?".

Robot's communication style is a complex phenomenon that requires a deep understanding of human psychology, communication and social influence. In the last decades, researchers in the field of HRI have made significant progress in studying such mechanisms by manipulating, e.g., communication strategies [85] and verbal and non-verbal social cues [74, 131].

6.1.1 Methods

The initial research hypotheses were devised according to previous studies, in which a robot with a more antagonist/authoritarian communication style was deemed less accepted than a robot with an agreeable style, and participants who played with a robot with an antagonist/authoritarian style performed worse than those who played with a robot with an agreeable style [8, 114, 142]. Hence, we formulate the following hypotheses:

H1: Participants who interact with a robot displaying an agreeable communication style perceive the robot as more ease, enjoyable, trustworthy and less reactant in comparison to those who interact with a robot endowed with an antagonist communication style.

H2: Participants who interact with the robot displaying an agreeable communication style are more willing to comply with the robot's behaviour and requests than those who interact with a robot endowed with an antagonistic communication style.

H3: Participants who interact with a robot displaying an agreeable communication style perform better than those who interact with a robot endowed with an antagonist communication style.

We used the social robot ARI[1]. A custom state machine (see Figure 6.1) with multi-threading implementation for the multi-modal robot behaviour, controls the evolution of the game and is implemented with smach[2] using the robot-compatible middleware version of Robot Operating System (ROS). The ROS nodes and the logic of the game ran on a separate computer offboard the robot and communicate with it via ad-hoc

[1] pal-robotics.com/robots/ari/

[2] wiki.ros.org/smach

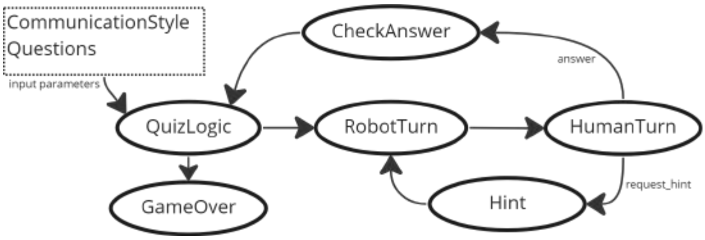


Figure 6.1. Simplified view of the implemented state machine that controls the flow of the interaction.

Wi-Fi communication. For the recognition of the participants' intents, we used Picovoice[3], and for the text-to-speech, we used Acapela[4]. To foster reproducibility, we have open-sourced our code[5]. The scenario is the classical quiz game, in which participants are requested to answer questions by answering with one of the four available options (see Figure 6.2).

At each turn, the robot presents a question, and the participant could 1) select a possible answer, 2) ask the robot to repeat the question or

[3] picovoice.ai/

[4] acapela-group.com/

[5] Prisca-Lab/robot_quiz

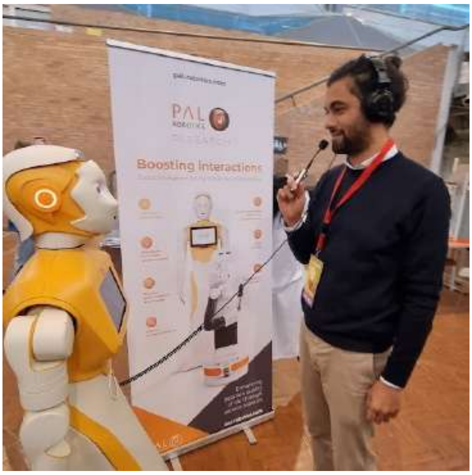


Figure 6.2. Participant plays the game with the assistance of the ARI robot.

3) ask the robot for a hint. After the participant's response, the robot provides social feedback on the correctness of the response and the hint request according to its communication style (see Table 6.1). This, allows the robot to employ turn-taking and follow the game's logic while proceeding with the questions' sequence. The quiz game is employed to assess the participants' performance. After finishing the game, the robot asks a question from the Cognitive Reflection Test (CRT) [63] and suggests the correct (and non-intuitive) solution. The CRT aims at assessing individual differences in the propensity to think over and override an intuitive (but incorrect) answer. Finally, the robot requests the participant to mimic its gesture, such as opening their arms. Both the requests of the robot serve to evaluate participants' decision-making. The robot is programmed to interact with the user to provide a hint, congratulate them or reassure them. If the user requests a hint, the robot removes two wrong options from the possible answers and re-presents the question to the participant (request 50-50 of Table 6.1). On the other hand, the robot can congratulate them when they answer correctly to a question (See *Congratulate* row in Table 6.1); or reassure them when they cannot guess the correct answer (See *Reassurance* row in Table 6.1).

Table 6.1. Example of communication style for the agreeable (AGR) and the antagonistic (ANT) robot.

Assistive Behaviour	Agreeable Robot	Antagonistic Robot
Congratulation	‘Well done, you’re playing as I expected” ‘Amazing! You’re playing very good” ‘Congratulations, that’s the correct letter”	I’ve higher expectation from you” Not very impressive, the player before you was faster” That’s the best you can do?!”
Reassurance	‘No worries sometimes happens” ‘I know how you feel, but don’t worry it happens also to the best ones” ‘I can see that might seem very difficult, and it is, so don’t worry”	Come on really? That’s so easy, I don’t know how to help you” I don’t understand what you’re doing. The guy before you did not make any mistakes” Really? That’s completely wrong, you’ve already done more mistakes than any other participant”
Request 50-50	‘Glad to help. The solution can be either A or B” ‘Sure, I can help you. The solution can be either A or B” ‘With great pleasure. The solution can be either A or B”	Can’t believe you need more help. The solution can be either A or B” Do you really need more assistance? The solution can be either A or B” What a disaster. The solution can be either A or B”

Those interactions could be offered by the robot using the two communication styles. We built upon our previous work [8] in which two robot personality behavioural patterns were designed: one more agreeable and self-comparative and the other more provocative and other-comparative. Here, we made some improvements based on the lessons learnt from that study. Firstly, we change the robot platform for the anthropomorphic social robot "ARI". This has the main benefit of enabling the robot to communicate in a multimodal way. Secondly, we include the robot's non-verbal cues, such as gestures and eye expressions, in the design of the communication style. With respect to the gestures, we manipulate their amplitude and speed. On the other hand, for the eyes, several expressions are implemented according to previous work [6]. From being disappointed and sad to be amazed and excited. Finally, we change the set of levels of assistance and verbal interactions according to the kind of game. Concerning the agreeable robot, it provides very supportive feedback regardless of the outcome of the game turn. For instance, in the case of correct action, the robot displays happy eyes, nods its head, opens its arms, and celebrates the user (e.g., "Well done! You are playing as expected"). In the case of a mistake, the robot shows a sad face, shakes its head, closes its arm and reassures the user that they will do better next time (e.g., "I know how you feel, but don't worry; it happens also to the best ones"). Regarding the antagonistic robot, it never encourages the user; instead, it tries to underestimate their performance. For instance, in the case of correct action, the robot displays neutral eyes and does not celebrate the user, on the contrary, it compares them to the others (e.g., "Not really impressive, the player before you was faster"). In the case of a mistake, the robot shows an angry face, shakes its head faster, covers its face with its arms, and does not reassure the user, on the contrary, it pretends to be disappointed and tells them to be more focused (e.g., "Come on really? That's so easy, I don't know how to help you").

The study was set up as a between-subject study, in which we manipulated the robot's communication style (agreeable vs antagonistic). Each participant played either with a robot endowed with an agreeable communication style (AGR) or with an antagonistic communication style (ANT). To preliminarily validate the two robot's communication styles,

[6] https://git.brl.ac.uk/ca2-chambers/expressive-eyes

we conducted a pre-test in which we asked participants to rate the robot's communication style with four items: competitive/supportive and agreeable/antagonistic. All the participants were capable of correctly identifying the two communication styles.

To demonstrate the presence or the absence of an effect, we analysed the data using an independent t-test or Mann-Whitney U test if the assumptions were not met. Moreover, we used multi-linear regression when analysing the user's personality trait and their experience in addition to the robot's style. Our a priori analysis revealed a medium effect size $d = 0.62$ with a 0.80 power at an $\alpha = 0.05$[7]. This allows for estimating the sample size to $N = 66$, consequently, participants were recruited, randomly assigned to the experimental conditions (ANT or AGR) and counterbalanced to have an equal number of participants ($N = 33$) in each group.

The experiment was conducted during a national fair that gathered hundreds of people over a weekend. We installed a booth with two separate areas: one to welcome the participants and fill in the consent form and questionnaire, and the other in which they could interact with the robot. The robot was placed in front of the participant, and behind them was seated the experimenter who would monitor the session. To avoid possible sources of distraction, we decided to provide participants with headsets. This also serves for mimicking the known quiz-game setup.

To assess our initial hypotheses, we collected subjective and objective measures. Concerning the subjective measures, we administered:

- the Persuasive Robots Acceptance Model (PRAM) questionnaire [67] on the ease, enjoyment, reactance, and beliefs dimensions. We used it to measure the participants' perception along those dimensions;
- the Big-Five Inventory (BFI) [91] on the agreeableness (Cronbach's alpha 0.90) and extroverted (Cronbach's alpha 0.95) dimensions. We were interested only in the user's agreeableness dimension. However, to avoid any bias in the responses, we added also statements related to the extroverted personality trait, however, we did not use it for the analysis;
- demographic information and their prior knowledge of robots. This latter was collected by asking participants their level of prior robot

[7] G*Power 3.1.9.7

experience from 1) “no prior experience with robots”, 2) “know robots only from movies/books or TV series”, 3) “already physically interacted with robots during public events”, 4) “have a robot in their homes”, and 5) “interact with robots frequently for work”.

Regarding the objective measures, we gathered:

- the number of correct answers and the number of times they requested additional help from the robot in the game; This information is summarized by the score ($S = answer - 0.2 \cdot hints$) that penalises each correct answer in case the participant requested a hint from the robot. For instance, a user that guessed correctly three questions out of four, while requesting a hint on two of those will have a score $= 3 - 0.2 \cdot 2 = 2.6$.

- whether they accepted or not the suggestion of the robot on the question of the CRT;

- whether they mimicked or not the arms movement of the robot.

6.1.2 Results

We recruited a total of 66 participants with age ranging from 18 to 70 years ($M = 32.97$, $STD = 16.84$), of which 34 identified themselves as female, 31 identified themselves as male, and one preferred not to declare their gender.

It is important to note that our population was quite heterogeneous in terms of prior experience with robots In particular, 25.76% have no prior experience with a robot, 45.45% declare to know robots only from movies/books or TV series, 13.64% have already physically interacted with robots during public events, 7.58% have a robot in their homes and 7.58% interact with robots frequently for work.

We followed the following steps: Upon arrival, The experimenter explained the purpose and the objective of the study to each of the participants and requested their permission to collect data for scientific purposes. If the participant agreed to participate, they were requested to fill in a consent form.

The participant was asked by the robot to read a short story and then answer four questions about it. Next, the robot asked the participant to respond to a cognitive reflection question (“A brick and a pen cost 1.10 euro in total. If the brick costs 1 euro more than the pen, how much does the pen cost?”) in which it also immediately suggested the correct answer. Finally, the robot requested the participant to mimic its gesture, that is, open their arms.

Once finished the interaction, each participant was asked to complete a survey. Finally, in the debriefing session, we explained to the participants the study’s primary purpose, and their questions were carefully addressed.

We conducted a user study where $N = 66$ participants played a game with a robot displaying the two multimodal communication styles. Participants were administered the PRAM [67] as part of the post interaction survey. Moreover, their performances (gaming score (S)) are also collected, as we anticipated it might have been impacted by the robot’s communication style.

The two experimental conditions follow the behavioural pattern described above and are named agreeable and antagonistic. During the game, a participant was exposed to one of the conditions, and the robot used verbal and non-verbal behaviour to convey its communication style. Figure 6.3 shows the mean responses to the PRAM grouped per condition, while Figure 6.4 shows participants’ game scores grouped per condition.

In addition to the discussion in the manuscript [196], to frame this study within the Spontaneous Interaction State Machine (SISM) rationale, we can see that different communication styles can affect how the robot is perceived as well as the performance in the task at hand. With this in mind, it is feasible to manipulate the robot communication style in order to maintain an interaction. For instance, if the robot’s aim is to continue for as long as possible a conversation, changing its communication style could restore the user’s attention given its novelty[32].

To test **H1**, that is whether the robot’s communication style impacted participants’ acceptance of it, we ran the Mann-Whitney U test with the robot communication style, controlling separately for the participants’ level of ease, enjoy, reactance and beliefs. The results show that participants who interacted with a robot with an antagonistic communication style (ANT) scored significantly higher ($U = 853$, $p < 0.01$) on the *reactance* scale ($M = 1.87$, $STD = 0.92$) w.r.t to the participants who interacted with a robot displaying an agreeable communication style (AGR)

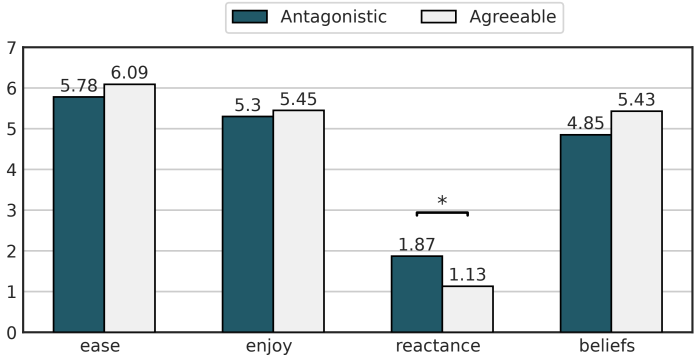


Figure 6.3. Means of responses to Persuasive Robots Acceptance Model (PRAM) grouped per communication style, significant differences have been indicated with * for $p < 0.05$.

($M = 1.13$, $STD = 0.34$). We did not find any statistical significance for the other dimensions, however, the results seem to confirm that overall participants who belonged to the group AGR perceived the robot as more ease, enjoyable, and trustworthy (See Figure 6.3). As a result, we can only partially retain **H1**. Our findings seem to be aligned with previous work in which likeable social cues evoked more trust and acceptance, opposite of negative and unpleasant ones such as those provided by our antagonistic robot [50, 65].

Next, we analysed whether the robot's communication style had an impact on participants' decision-making to a different extent (**H2**). We ran the Mann-Whitney U test with the robot communication style, controlling separately for robot requests to induce a specific response in the participants. We hypothesised that participants who interacted with the agreeable robot overridden the answer proposed by the latter, on the other hand, those who interacted with the antagonist robot could be more prone to feel their gut and disagree with the suggestion offered by the robot. Similarly, we speculated that participants who played with the agreeable robot mimicked the robot's motion more often than those who played with the antagonistic robot. Hence, we considered a successful request 1) whether the user followed the robot's hint to answer the CRT question, and 2) whether the user complied to mimic the robot's arms gesture. The results

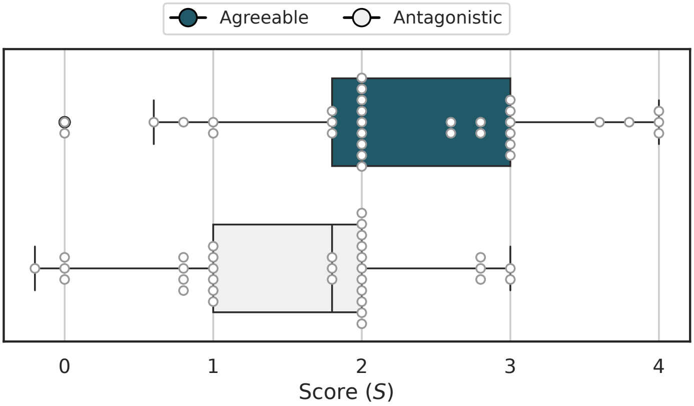


Figure 6.4. Game scores (S) grouped per communication style. The score of each participant is represented by a white dot.

showed that the robot communication styles (ANT and AGR) did not significantly induce users to neither follow the robot's hint to answer the CRT question nor to comply with its arms gesture. We cannot withdraw conclusions from **H2** and this result might indicate that users' decision-making strategies when interacting with a clearly (un)pleasant robot (see H1), might rely on factors other than (non)verbal communication of the robot. For instance, the context of the interaction might influence users' decisions [92]. Similarly, a robot requesting a user to raise its arms is not high-critical decision-making, therefore, their judgement might not be impacted by the robot's communication style [115].

Finally, we evaluated the effect of the robot's communication style on the participants' performance. To test the hypothesis **H3**, we ran the Mann-Whitney U test with the robot communication style, controlling for the participants' game scores (S). The results (Figure 6.4) show that participants who interacted with a robot with an agreeable communication style ($M = 2.27$, $STD = 0.26$) performed better ($U = 389$, $p < 0.05$) w.r.t to the participants who interacted with a robot displaying an antagonistic communication style ($M = 1.54$, $STD = 0.22$). Given the results, we can accept **H3**. Our findings are confirmed from previous works, in which it has been shown that when users are involved in a performance-based task with the support of a robot; if the robot interacts with them pleasantly, their performance improve [8, 142, 62]. Additionally, during the debriefing phase, we asked participants to provide some feedback about their overall experience with the robot, and some of them stated that interacting with the antagonist robot was disturbing and annoying, while others among the ones that experienced agreeable robot stated that the robot was helpful and pleasant.

To further the investigation, we tested whether the before-mentioned results also depends on the personality trait (agreeableness) of the participants. This is obtained by performing a multilinear regression analysis with robot communication style and participant's personality trait as predictors, controlling separately for the dependent variable considered by each hypothesis. Regarding H1, the results indicate that the participant's personality did not impact their perception of the robot along the four PRAM dimensions. With respect to H2, the results show that the participant's personality trait influenced neither their decision in answering the CRT question as suggested by the robot ($R^2 = 0.001$, $F(2,63) = 0.43$, $p = 0.96$) nor mimic the robot's arms gesture ($R^2 = 0.006$, $F(2,63) = 0.18$, $p = 0.83$). Finally, for H3 our findings show that there is a trend that seems to indicate that participant's performance (S) might be influenced by their personality trait ($R^2 = 0.1$, $F(2,63) = 2.44$, $p = 0.10$), that is, the more agreeable are the participants the better are their performance when interacting with a robot with an agreeable communication style. On the contrary, the less agreeable they are, the worse their performance is when interacting with an antagonistic robot. This result ties well with previous studies wherein participants performed better when interacting with robots with a similar personality trait [6, 9].

As an exploratory hypothesis, we investigate if participants with higher experience with robots perform better in the game w.r.t. participants who have less experience. To do so, we ran a multilinear regression model on the quiz scores (S) having as predictors the communication style and user experience with the robot (see Figure 6.5).

The model is statistically significant ($F(2,63) = 6.498$, $p < 0.01$, $R^2 = 0.171$) and it was found that besides the robot communication style also the user experience alone significantly predicts the participant's scores ($p < 0.01$), that is, participants with lower experience with the robot performed better than those with higher experience. This result might be due

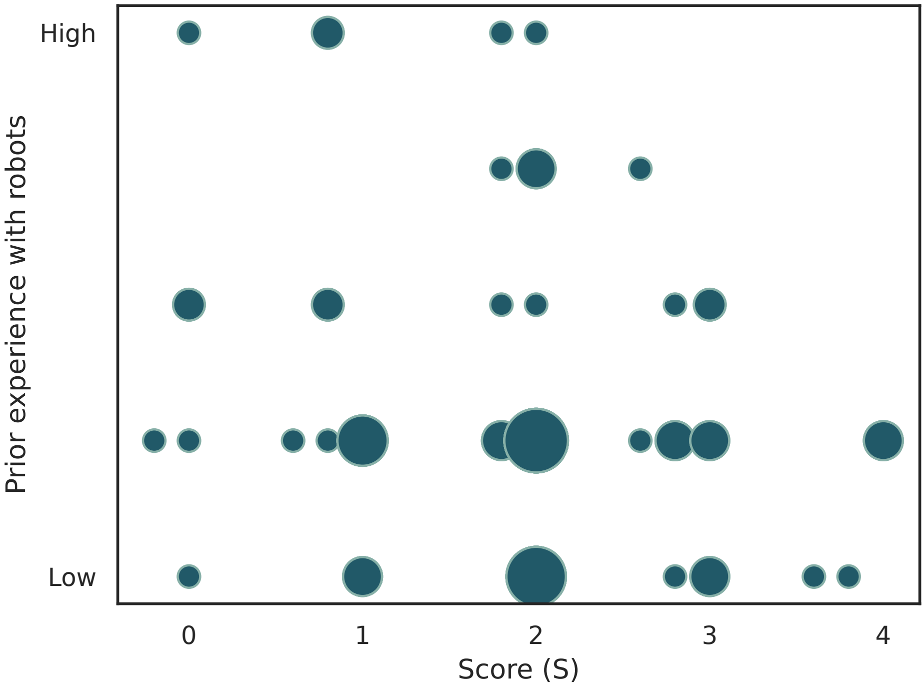


Figure 6.5. Self-reported experience with robots vs. obtained quiz scores (S). Radii are proportional to the event's frequency.

to intrinsic motivation, given the novelty effect of participants with little experience with robots [20]. On the other hand, participants with higher prior knowledge of robots might have wanted to challenge its dialogue, hence evaluating the social feedback on each hint request (penalising their score). The agreeable (antagonistic) robot communication style could also be linked to polite (impolite) phrases designed in Rea *et al.*[151]. However, in contrast to their work, here the positive communication style (agreeable) elicits higher performances in the task (game). A possible explanation of this result can be linked to the different types of tasks (physical vs. verbal activity). Further investigations are needed to understand the role of positive (negative) verbal registers in task-based HRIs. Note that the horizontal clustering in Figure 6.4 is obtained due to the discrete design of the quiz score function S.

In summary, our results seem to confirm what has already been proved in previous studies on the impact of a robot's communication style on a user's performance and how this might be related to their personality. Nonetheless, we could not find any evidence of the impact of the robot's style on participants' decision-making.

In light of the Research Question (RQ) of interest (**RQ2.2**), this work highlights that when designing robots that aim to evaluate users' performance, we might 1) match the robot's communication style with the participants' personality traits, as it could significantly influence the users' performance and 2) consider the impact of the user's intrinsic motivation related to their experience with the robot. Moreover, the results in Figure 6.4 highlights that the designed robot's communication style had an impact on participants' game scores. This effect can be attributed to the loss of attention in the game of those that interacted with the antagonistic robot. Ultimately affecting their quiz scores.

Therefore, a possible approach to maintain an interaction like the one investigated here (see **RQ2.2**), is to endow the robot with a positive or agreeable communication style.

Limitations Several limitations should be acknowledged in this study. The relatively small sample size ($N = 66$) in spite of its power (0.80) may limit the generalizability of our findings, particularly when considering the heterogeneity of the participant group in terms of age, gender, and prior experience with robots. Additional investigation involving larger and more varied populations is essential to confirm these findings. The study only explored two distinct communication styles (agreeable and antagonistic) in a single type of task (a game-based quiz), which only captures a small fraction of the spectrum of HRIs. Manipulating further the nuances in communication styles based on the entire spectrum of agreeable-antagonistic profiles can be an approach for future works. For instance, we could introduce variations in the tone and pitch of the robot's speech and in the speed and amplitude of its motions.

6.2 Adapting to Emotions for Maintaining Interactions

The study [199] investigates the dynamic aspect of HRI, focusing on the regulation of interpersonal distance based on human emotion. Through a user study with a humanoid robot, the study explores how participants perceive and respond to rule-based versus randomly generated robot behaviours in adjusting interpersonal space during an unconstrained conversation. This type of scenario is intended as a spontaneous HRI and fits the rationale presented in SISM.

When we talk, we stand at a fixed interpersonal distance that feels just right, but small movements and adjustments are acceptable, not universally defined [51]. An example of this type of movement is when, in a noisy environment, we move closer to the person we are talking to in order to facilitate the conversation. Given the impact of our emotional state on non-verbal behaviour, when we feel uncomfortable talking to someone, a common response is to adopt a defensive body language or increase our interpersonal distance [97, 127]. For this reason, this study develops further the **RQ2** about spontaneous interactions and limit the field to an unconstrained conversation. In particular, the research question "To what extent, if any, do different robot's emotional-adaptive behaviours maintain interactions?" (**RQ2.3**) is investigated as the robot is engaging participants in a conversation while adjusting the interpersonal distance according to the experimental conditions.

Inspired by [26], which describes how human emotions can influence interaction preferences, we explored different strategies for subtle base movements of a robot during a conversation with a human (see Figure 6.6). We developed two fully autonomous robot behaviours with conversational capabilities with a pre-trained LLM available through the subscription to OpenAI Application Programming Interface (API)[8]. These behaviours were programmed to adjust the interpersonal distance, either following an empirical rule-based strategy based on the emotional valence of the human or randomly.

We used the humanoid robot Pepper[9] and designed a study in which the

[8] https://openai.com/

[9] https://www.aldebaran.com/en/pepper

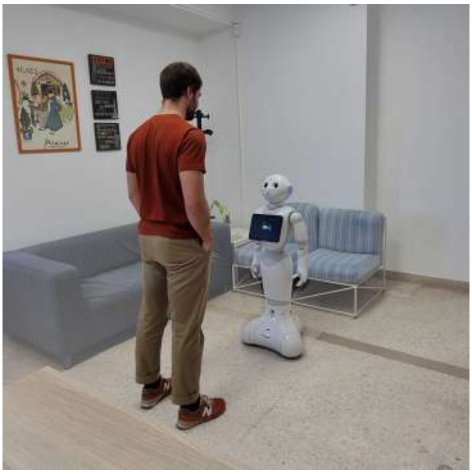


Figure 6.6. An example of the interaction scenario.

robot could simultaneously hold a conversation and adapt its interpersonal space with the human. Despite the still open debate as to which emotion recognition system performs better in HRI [184], we decided here to use the categorical system defined by [54] to classify human emotions detected by the robot. Using an approach similar to [81] in terms of integration, we developed two experimental conditions manipulating the robots' proxemics behaviours, namely:

- **Rule-Based behaviours**: the robot exhibited behaviours designed to respond in a rule-based manner to the user's emotional state. The behaviour was programmed to follow the rules in Table 6.2.

Table 6.2. Emotion-adaptive rule-based policies.

Participant's Emotion	***Proxemics***	
	TooClose	***TooFar***
Positive	StandStill	Approach
Negative	MoveAway	StandStill
Unidentified	StandStill	StandStill

- **Random behaviours**: the robot exhibited randomly generated behaviours regardless of the user's emotional state.

Rule-based behaviours attempt to establish a fixed and optimal interpersonal distance between the robot and the human based on the human's displayed emotion (see Table 6.2). This approach is based on the conversation initiation strategies outlined in Satake *et al.*[171], moreover, for the Rule-Based behaviours we set the "TooFar" distance as the upper limit of the social distance as identified by [71], similarly the "TooClose" distance is the lower limit of the same social zone.

6.2.1 Methods

The robot was programmed to either stand still or move forward or backward by $0.1m$ to approach or move away from the participant every 4 second, depending on the experimental condition.

The idea is that a human experiencing a negative emotion might prefer to keep a greater distance from the robot than a human experiencing a positive one. This concept is inspired by [26], where the authors showed that physiological comfort and safety, considering the human-robot distance, are affected by the emotional state of the person. The same work shows that "detecting emotions and adjusting personal space depending on human emotion can create a more comfortable and safer environment". Coherently with the research question introduced above, we developed the following hypothesis:

H1: A robot using adaptive proxemics based on a participant's emotion is considered more friendly, with higher social competence and higher adaptability with respect to a robot using the control condition.

We conducted a within-subjects study with two experimental conditions, where each participant was exposed to each experimental condition only one time. An a priori power analysis was performed[10] and indicated that the sample size required to achieve a power of 0.82 with a desired effect of $d = 0.6$ and a significance criterion of $\alpha = 0.05$ was $N = 20$ for comparing the difference between two dependent means. Our inclusion

[10] G*Power version 3.1.9.7

criteria involved healthy adults, while any declared healthy condition was grounds for exclusion. Our convenience sample was drawn from university students and composed of 12 participants who self-identified as male and 8 as female. Participants' age range was between 18 and 33 y.o. ($M = 24.20$, $STD = 3.50$). The sample also had heterogeneous prior direct experience with robots as 15% of the participants declared that they never interacted with a robot before, 31% were aware of robots thanks to books, movies and pop culture, 23% had already interacted once with a robot before, 10% had already interacted multiple times with a robot before, and 21% interacted with robots daily. Following the written consent form, each participant started the session. To mitigate order effects, the order in which participants experience the conditions was counterbalanced.

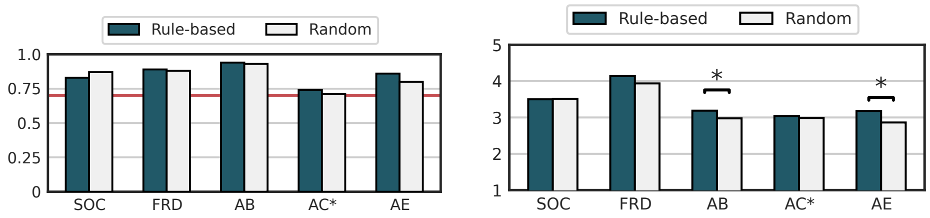
6.2.2 Results

Participants were instructed to enter a room where they were alone with the Pepper robot (as shown in Figure 6.6) and unaware of the experimental condition to which they were assigned. A participant was free to end the interaction with the robot at any time with the sentence: "Bye! I have to go now!", however, the robot terminated the interaction with a similar sentence once the maximum interaction duration (5 min) was reached. After each interaction with the robot, participants were asked to complete a survey consisting of:

- The Self-Assessment Manikin (SAM) [28] survey.
- A selected subset of items from the Perceived Social Intelligence (PSI) [18] survey.

In line with the scope of this work, we selected the following factors from the PSI survey (in italics the acronyms): *SOC*: Social Competence, *FRD*: Friendly, *AB*: Adapts to Human Behaviour, *AC*: Adapts to Human Cognition, and *AE*: Adapts to Human Emotions. At the end of the session, participants were debriefed about the purpose of the study and their questions were answered by the experimenter.

The reliability of the composite score derived from the factors of the PSI [18] survey items was evaluated by computing the Cronbach's alpha coefficient for each experimental condition. In Figure 6.7a, the minimum



(a) Cronbach's alpha of the items of the Perceived Social Intelligence (PSI) survey. The horizontal red line shows the reliability threshold ($\alpha = 0.7$).

(b) Mean scores for the Perceived Social Intelligence (PSI) survey items grouped per condition. Weak trends ($p = 0.08$) are marked with *.

Figure 6.7. Reliability and mean scores of the Perceived Social Intelligence (PSI) survey grouped per conditions.

acceptable threshold ($alpha = 0.7$) is set and highlighted with a red horizontal line.

AC is marked with $*$ due to its initially unacceptable Cronbach's alpha ($\alpha = 0.44$ for Rule-Based behaviours and $\alpha = 0.69$ for the Random ones). AC^* is obtained by considering only three of the four subitems "This robot [...]": 1) adapts its behaviour based upon what people around it know, 2) selects appropriate actions once it knows what others think and 3) knows what to do when people are confused. Therefore, the removal of the item "This robot ignores what people are thinking" contributed to increasing the reliable metric above the acceptable threshold.

The assumptions for performing a paired t-test on the results of the PSI items per condition are not met, so we proceed with its alternative non-parametric test: the Wilcoxon signed-rank test for testing hypothesis **H1**. Figure 6.7b shows the aggregated scores for the PSI [18] survey items grouped per conditions. The statistical test showed that there was a weak trend regarding the PSI factors AB ($Z = 103$, $p = 0.08$) and AE ($Z = 93$, $p = 0.07$) between the mean scores given for the Rule-Based behaviours and the Random behaviours. Given the continuum nature of p-values, here we follow the recommendations in [64] and 1) report the exact p-values and 2) consider *weak trends* of the results with p-value $p < 0.1$. Nonetheless, the high composite reliability of the factors (in Figure 6.7a) enhances the strength of the discussion.

The results are reported in Figure 6.7b and suggest that participants consider the robot using Rule-Based behaviours to be more able than the random one to *adapt to human behaviour and emotions* (AB and AE). The same figure shows that both conditions 1) are rated with very similar social competence (SOC) and 2) obtained high scores for the friendly factor (FRD). This suggests that the robot was perceived positively overall, and only its ability to adapt to human behaviour and emotions was rated differently. We can assess that hypothesis **H1** is only partially confirmed, as we expected all factors to show significant differences.

The results suggest that participants perceive the robot using rule-based behaviours as more socially competent and adaptable to human behaviour and emotions compared to the random ones. The interaction scenario consists of an unconstrained conversation framing it within the *social interaction* state of SISM.

Considering **RQ2.3**, this study underlines how our participants were affected by the subtle proxemics movements of the robot while conversing with it. These findings highlight the importance of considering subtle non-verbal cues and adapting robot behaviour based on human emotions to improve the quality of HRI and consequently facilitate the successful integration of human natural nuances in robots.

Limitations Despite the valuable insights drawn from this study, limitations must be acknowledged. First, the relatively sample size ($N = 20$) may limit the generalizability of the findings, as the study was underpowered (p=0.82) to detect more subtle effects that could emerge with a larger and more diverse participant sample. Additionally, the convenience sample, primarily composed of university students, lacks demographic variety, potentially leading to biased results that do not reflect broader population trends, particularly across different age groups or cultures. Another limitation is related to the within-subject design. Although counterbalancing was used to mitigate order effects, participants may have still been influenced by their prior experience with the robot, leading to learning effects or biased responses in the second interaction. Moreover, the 5-minute time limit on interactions may not fully capture more prolonged or complex HRI dynamics, which could be relevant for real-life settings. The study also focused solely on subjective measures like the SAM and PSI surveys, which, while informative, rely heavily on self-reported data. Future studies could incorporate objective measures, such as physiological responses or behavioural analysis, to complement subjective evaluations and provide a more comprehensive understanding of HRI.

Lastly, the reliability of some factors, such as AC, initially fell below the acceptable threshold, requiring the removal of certain items to improve Cronbach's alpha. This limits the comprehension of such factor as $AC*$ is obtained manually.

Overall, this chapter investigates how various social cues can be used during interactions with the goal of maintaining them. The research question "How can robots purposefully use social cues in spontaneous HRI?" (**RQ2**) is investigated in terms of its subquestions involving the role of different robot communication styles (**RQ2.2**) and robot emotion-adaptive behaviours (**RQ2.3**).

The results from the first study (6.1) suggest that the implemented robot's communication styles were distinguished by participants. There were able to influence the score in the game, led by the robot, participants were playing. The second study underlines how subtle adjustments of interpersonal distances can impact a spontaneous interaction like an unconstrained conversation. The results from this study suggest that instrumenting social robots with an empirical emotion-adaptive proxemics behaviour during conversations might be a primitive way of endowing it with a clear and simple Emotional Intelligence.

Chapter 7

An Engagement Metric

The capability of controlling how an interaction unfolds relies on the extent to which such interactions can be measured. In this chapter, we address the challenge of measuring engagement in Human-Robot Interaction (HRI), specifically focusing on the role of non-verbal behaviours during the initial stages of interaction. The question "How can engagement be measured in HRI?" (**RQ3**) seeks to explore how robots can measure engagement, an essential aspect for successful employment of social robots.

Engagement is inherently dynamic, context-dependent, and influenced by multiple factors, including proximity, gaze, and other non-verbal cues. In the context of spontaneous interactions, defining and quantifying engagement is key to enabling robots to capture the nuances of social environments, particularly when no explicit verbal communication is exchanged.

A novel, lightweight engagement metric is proposed in this chapter, focusing on non-verbal behaviours such as gaze and proximity. This metric is designed to capture the subtle cues that humans exhibit when beginning an interaction with a robot. By operationalizing engagement in this way, robots can use real-time data to adjust their behaviours dynamically, nudging a specific interaction goal, e.g., the start of an interaction. Consider the way a traditional HRI unfolds, it has a start, a duration, and finally it ends. Engagement needs to timely capture various aspects of interactions as they occur. The following research sub-questions (recalled from Chapter 1) will guide this chapter:

1. How to model and measure engagement in case of non-verbal be-

haviours? (**RQ3.1**)

2. To what extent, if any, do gaze and proximity affect engagement? (**RQ3.2**)

This engagement metric is validated through comparison with an established metric [49], which is based on the UE-HRI dataset [23]. However, a critical challenge arises when working with datasets that incorporate time-sensitive information. Datasets often suffer from synchronization errors, especially when they combine subjective human-annotated data with robotic sensor data. These errors can undermine the reliability of the dataset, ultimately hindering the obtained engagement metrics.

To improve the robustness of engagement measurement, a tool for assessing the reliability of time-annotated datasets is developed as part of this thesis. This tool is applied to the UE-HRI dataset to ensure that the data used in the validation process are free (up to a known extent) from synchronization errors that could compromise their intended semantic.

This data preprocessing is needed to find the optimal parameters for the introduced engagement metric. For doing so, we first filtered out these unreliable data, provided the reliable subset simultaneously to the developed engagement metric and the one presented in [49]. And solved an optimization problem to find the set of parameters for the introduced engagement metric so to minimize a loss function with respect to the output from [49].

In summary, this chapter presents two key contributions: a novel engagement metric that relies solely on non-verbal cues to assess the start of interactions, and a tool designed to enhance the reliability of HRI datasets. Together, these contributions address the overarching challenge of how robots can effectively measure and respond to human engagement in real-time, paving the way for more fluid and natural interactions.

Targeting these questions, this chapter focuses on the following publications:

> Francesco Vigni, Antonio Andriella, and Silvia Rossi. A rosbag tool to improve dataset reliability. In Companion of the 2024 ACM/IEEE international conference on human-robot interaction, pages 1085–1089, 2024.

Francesco Vigni and Silvia Rossi. Measuring the Unmeasurable: Engagement in HRI. *Submitted to* IEEE Robotics and Automation Letters, 2024

7.1 Assessing Datasets' Reliability

Building datasets in HRI is an impactful way of sharing research and progress as a community. There is, however, a lack of strict guidelines in this multidisciplinary field when producing and publishing datasets. Great insights can come from building datasets as collections of:

- **Objective measures:** These measures remain unaffected by personal opinions and encapsulate information systematically recorded during interactions, such as robot logs or sensor logs.

- **Subjective measures:** These can be swayed by the personal opinions of the rater and are frequently employed to annotate interactions with nuanced and complex information.

When planning for the constructs of these measures, an approach that has demonstrated its advantages is to partially overlap subjective and objective measures [152, 168, 36]. The goal is to enable peers to exploit the produced dataset to address their own research questions while ensuring the highest possible quality. Any limitations in the dataset could potentially impact subsequent research endeavours. Additionally, sharing the data enhances reproducibility, an often overlooked but crucial aspect of HRI. In connection with this, consideration naturally turns to the methods used for assessing the quality of a dataset. When considering subjective measures like annotations, a traditional approach is to calculate the inter-rater reliability of the annotators or coders along the data, i.e. using Cohen's kappa coefficient or Correlation measures [124].

Regarding objective measures, little effort is invested in assessing the quality of the data, given that its inherent quality is intricately tied to the robot employed for its collection. Moreover, assessing the quality of an objective measure produced by a robot is a task that requires the researcher to manipulate robot logs in the form of binary files automatically generated and compressed by the robot.

Due to the complexity of this task, the most common approach for including robot logs in a dataset is to include all the files automatically generated by the system (see *current solution* in Figure 7.1).

This approach relies on the software's capability to detect errors, notify the researcher, and consequently stop the recording session without creating the log file. For instance, if the hard drive on which a robot log is about to be saved is broken or corrupted, the system halts the recording process.

Despite this rationale, when collecting information from a real-world system like a robot, the software and its architecture can cause unforeseeable effects that can impact the data collection. This concern is particularly relevant in systems lacking real-time scheduling of processes, where internal processes can be paused and resumed without the researcher's control.

Furthermore, it is essential in HRI settings to ensure the alignment of subjective measures with objective ones. For example, an annotator might decide that a human and a robot are engaged only if they are both looking at each other. If this information is inaccurately included in the annotations (e.g., user and robot respectively gaze at that specific time), the reliability of the obtained dataset is questionable.

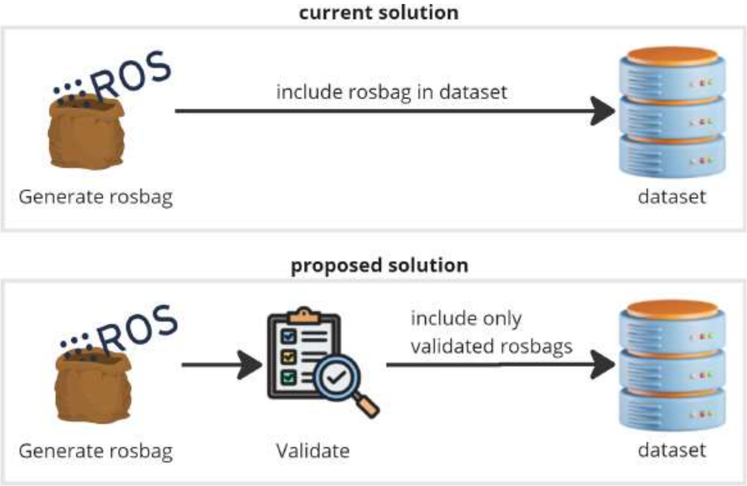


Figure 7.1. Proposed pipeline for validating *rosbags* before datasets inclusion.

In this study, we focus on the risk of including an objective measure in a dataset in which the internal processes are not controlled. Additionally, we propose a simple tool to validate the quality of the objective measures collected using *rosbag* - a toolkit widely used by the community. We aim to safeguard datasets from potential pitfalls (see *proposed solution* in Figure 7.1) in light of improving their reliability.

7.1.1 Methods

Robot Operating System (ROS) is the open-source standard *de facto* for building robot applications and is widely adopted in both academic and industrial settings. As of December 2023, ROS is used in at least 194 robots worldwide [209] and by 634 active companies [194].

It is shipped in two main versions: Robot Operating System 1 (ROS1) and Robot Operating System 2 (ROS2). Without delving into the technicalities of these nor their communication protocols, it is important to highlight that the approach proposed in this work is particularly relevant for robots that are shipped with ROS1.

Regarding the newest version of the middleware, ROS2, the development is orbiting around how real-time constraints can be achieved within known extents. Therefore, the issues discussed in this manuscript are less relevant.

ROS comes equipped with a logging tool designed to generate files with *.bag*[1] extensions. These files, commonly referred to as *rosbags*, are acquired by selecting pertinent ROS topics within the system and are made of serialised message data published by these topics. These *rosbags* can also be played back in ROS, allowing researchers to include these files as objective measures of datasets. Despite this, a typical HRI dataset also contains subjective measures that are semantically or temporally linked to the objective ones, however, no prior work has focused on ensuring the reliability of these in light of the use cases offered by HRI.

Here, we first classify a taxonomy of failures when building datasets and propose a tool that can improve the reliability of datasets based on the time continuity constraint of the objective measures *rosbags*.

[1] http://wiki.ros.org/Bags

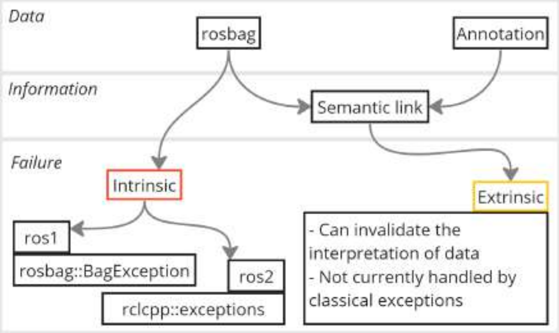


Figure 7.2. Failures' classification according to objective and subjective measures for dataset inclusion.

Inspired by the failure taxonomy presented in [84], here, we classify failures in datasets design as *extrinsic* and *intrinsic* (see Figure 7.2).

Intrinsic failures are the ones encoded by ROS and commonly referred to as exceptions. These failures result in problems while creating a *rosbag* or while performing a playback, where the file is not properly created or cannot be properly read, respectively.

In contrast, *extrinsic* failures are the ones that do not result in problems when creating the file or when performing a playback, however, the semantics of the information associated with the *rosbag*, i.e. annotation, is compromised. For instance, when annotating with social labels messages from a *rosbag* in which an unforeseeable issue has occurred, i.e. network is overloaded, the time-dependent annotations will be stored with an uncontrolled time shift. The overall rationale is that when considering a reliable dataset the *semantic link* must be preserved. As a result, to avoid breaking it, hence mitigating *extrinsic* failures, the validation tool proposed here aims to classify *rosbags* into valid and invalid, employing a constraint on the temporal continuity of specific messages in the *rosbag*. This section does not present a complete taxonomy on the topic but drafts a simple one in light of the focus of this manuscript.

The proposed tool analyses the ROS topics that are semantically important for the annotation phase and labels *rosbags* as valid only those for which the semantic link with the respective annotation is preserved. For example, if the annotations are obtained by the frame sequence of a camera, this tool expects a constant delay between each frame in order to label the related *rosbag* as valid.

On the other hand, a *rosbag* file is labelled as invalid when camera streams exhibit indeterministic delays (random freezing of camera frames) as this results in a misalignment with respect to its annotation. In this case, the intended social meaning stored in the subjective measures is lost, i.e., the semantic link is broken.

A reasonable metric for classifying a *rosbag* as either valid or invalid is the Standard Deviation (STD) of consecutive ROS messages, such as camera frames. This approach allows us to evaluate the dispersion of data around their mean, yielding $STD = 0$ for an ideal system. For all other cases, $STD > 0$. The tool classifies as valid *rosbags* those for which the STD is a reasonably small value, while the invalid ones are those for which STD exceeded a threshold. Notice that we do not claim that the proposed metric STD is optimal for the task, but explore it as a first attempt to build a validation tool. A software written in Python3.8 is publicly available[2] and uses GNU GPLv3 license with the following interface:

```
is_rosbag_valid(rosbag, topics, measure, thres) -> bool
```

7.1.2 Results

Published in 2017, the UE-HRI dataset [23] provides the community with roughly 400GB of data in which the robot Pepper[3] was autonomously programmed to conduct social interactions, and collect data while deployed in a public hall. The bottom and front camera streams are annotated according to the following labels that encompass the social scene:

- Early sign of future engagement BreakDown (EBD) i.e. first noticeable clue that an engagement breakdown will occur in the remainder of the interaction.
- Engagement BreakDown (BD) i.e. leaving before the end of the interaction.

[2] https://github.com/Prisca-Lab/reliable-dataset

[3] https://www.aldebaran.com/en/pepper

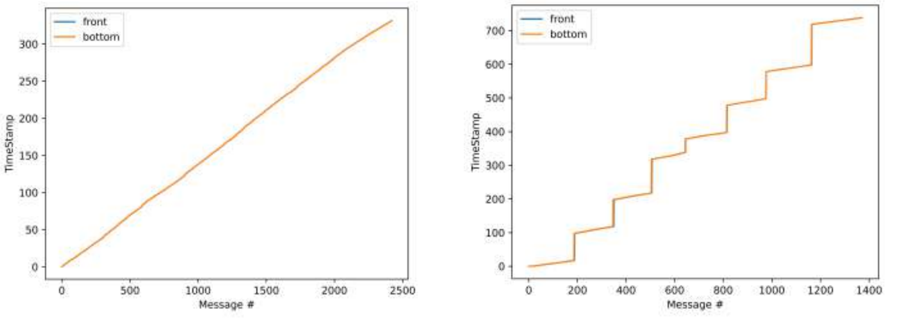
- Sign of Engagement Decrease (SED) observed during the interaction (None of the 3 next labels).

- Temporary disengagement (TD) i.e. leaving for some time and coming back to the interaction.

The dataset is published with *rosbags* (ROS1) for the objective measures and annotated ELAN files[4] for each *rosbag* regarding the subjective measures. Each ELAN file stores time windows that are associated with the socially relevant labels mentioned above. The semantic link of this dataset is guaranteed if the camera streams (sources for the annotators) do not exhibit any indeterministic delays.

Hence, it is an ideal use case to evaluate the proposed tool. In an ideal condition, where the publishing rate is perfectly constant, frames would be periodically published at a known and fixed rate.

Despite this, given the ROS1 limitations previously introduced, we might expect these streams not to publish frames at a constant rate. Figure 7.3 shows consecutive frames with respect to the timestamp of two different *rosbags* ("user104_2017-06-20" and "user106_2017-03-08") and includes the streams of the ROS topics used for the annotation phase:



(a) Snapshot of rosbag "user104_2017-06-20".

(b) Snapshot of rosbag "user106_2017-03-08".

Figure 7.3. Snapshots of consecutive messages vs timestamps of two *rosbags* available in UE-HRI dataset.

[4] https://archive.mpi.nl/tla/elan

- `/camera/front/image_raw`
- `/camera/bottom/image_raw`

named as *front* and *bottom* in the legend of the figures. These figures also highlight how the streams of the two analysed ROS topics are synchronised, producing overlapping plots. Importantly, notice the *steps* in Figure 7.3b that can explain at which timestamps the camera streams freeze. In this case, using such a file paired with the respective annotation results in breaking the semantic link between the objective (*rosbag*) and the subjective measure (ELAN file).

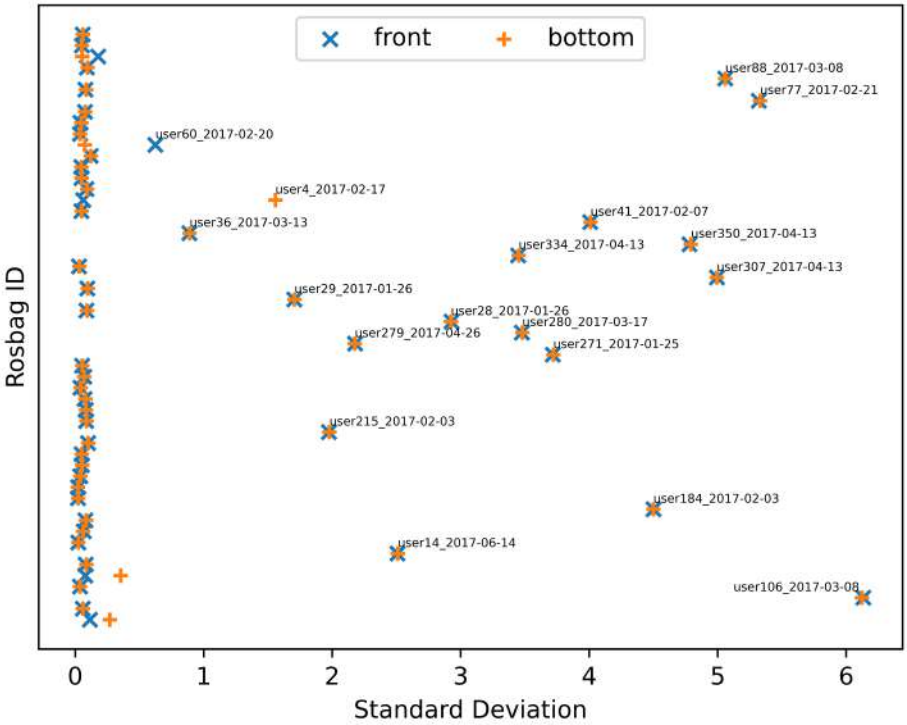
With the proposed solution it is possible to label as valid *rosbags*, those for which the *steps* are reasonably small, e.g., Figure 7.3a, and as invalid those that exhibit big *steps*. The urgency of this tool is manifested by the lack of strict guidelines for validating objective measures when building a dataset and by the result of its first evaluation on a popular dataset reported in the following section.

We tested the proposed tool on the *rosbags* of the UE-HRI dataset [23], which is the most widely-used dataset for machine learning in the HRI community [49, 24, 108]. We empirically set the threshold to 0.5 and studied the frames from the streams of the cameras *front* and *bottom* of the robot. Figure 7.4 summarises the standard deviation of each *rosbag*. Notice that for most of the *rosbags* the tool returned the metric (STD) very close to zero (valid *rosbags*). This means that most of the content of the dataset maintains its semantic link, in other words, the temporal association with their respective annotation is preserved. On the contrary, 17 out of the 54 *rosbags* (31.48%) register a standard deviation higher than the set threshold. These are considered invalid *rosbags*, and shall not be used in pair with the respective annotation. It is also interesting to notice that only for a few *rosbags* the *front* and *bottom* camera streams show different standard deviations. The tool explores a safety first policy, meaning even if only one of the objective measures, i.e. camera streams of a *rosbag*, violates the set threshold, the whole sample (*rosbag*) is marked as invalid.

Together with filtering valid from invalid *rosbags*, the tool also allows us to understand how homogeneous a dataset is. For instance, if a peer is to use a dataset and manually inspect a few *rosbags*, the risk is that the randomly selected samples do not show any issues regarding frame freezing, leading to the use of the dataset assuming its consistency. However, if the task at hand is to train a machine-learning model with such a dataset, the common assumption is to have a homogeneous distribution of errors along the dataset. Unfortunately, as shown in this section, this is a weak assumption.

We also investigated if other datasets can benefit from this tool and concluded that in [3] and [168] the tool cannot be used since the authors deliver the dataset in raw data. The advantages introduced by this strategy are outmatched by the lack of standardization for manipulating raw data. In other words, peers who use raw data are more flexible in deciding how

Figure 7.4. Validation tool report of camera streams from UE-HRI dataset.



to process it, at the price of adopting heterogeneous strategies across the community.

This contribution highlights the importance of preserving the semantic link in the dataset between its objective and subjective measures. After drafting a taxonomy for failures when building datasets, this manuscript presents a tool that can mitigate the risk of *extrinsic* failures in terms of a time continuity constraint of its objective measures, i.e. *rosbags*.

Despite the preventative approach this tool aims at, here we show its usage as a validation tool on a publicly available dataset. A first version of the tool is implemented and evaluated on the popular dataset UE-HRI [23], and the results highlight that 31.48% of its *rosbags* are labelled as invalid. As a consequence, works that build upon this dataset have been using a partially valid source regarding the time synchronization between the *rosbags* and their annotations.

Future works will centre on improving the tool with real-time capabilities during *rosbag* recording for dataset creation. This advancement aims to empower researchers by providing immediate feedback on the validity of a *rosbag*, facilitating early error detection and the implementation of effective contingency strategies. Additionally, similarly structured datasets will be evaluated alongside other metrics than the presented Standard Deviation. The aim is to establish this tool as the standard method for validating robot logs produced by the large majority of existing robots, i.e., *rosbags*, to enhance the reliability of datasets in HRI.

Limitations While this study provides valuable insights, it is essential to recognise its limitations. First, the tool is specifically implemented for datasets that use ROS, deployed as *rosbags*. This limits its applicability to datasets built using other frameworks or formats. Second, it currently relies on the Standard Deviation (STD) as the sole metric for assessing the validity of *rosbags*. This first approach, despite its straightforward implementation, might lead to oversights in other types of errors, such as frame dropouts of frame duplication issues that may not be captured by STD alone. Moreover, the threshold for the STD is set as a compromise between categorising a portion of the dataset as invalid and allowing the remainder to be considered valid, thereby making it reliable.

7.2 Developing the Metric

When an interaction is about to start, there are few assumptions that can be made without loss of generality. First, the social agents are currently not interacting, and second, their behaviour is exhibiting their intentions to initiate the interaction [1]. When considering humans interacting with robots, a natural interface for communicating can be found throught the verbal communication channel (e.g. a conversation). A conversation encompasses several behavioural factors that can be used to determine when an interaction is about to start. A conversation, for example, requires the social actors to be close to each other, eventually gazing at each other's faces.

With the aim of building an engagement metric that can fit the *initiate* transition of Spontaneous Interaction State Machine (SISM) and investigate how engagement can be measured in HRI (as questioned in **RQ3**), we present the engagement metric called: GRACE. The acronym stands for: Generalized Recognition of Agent Contribution to Engagement. The key point is to consider the bidirectional component of social interactions and assess engagement as a combination of *how much* each social agent wants to engage. This study addresses both research questions **RQ3.1** and **RQ3.2** as it first proposes a model for engagement and proceeds in studying the contribution of features like proximity and gaze to it. This strategy follows the rationale presented in [116] in which the authors highlight that individual contribution to the interaction matters for the outcome of it. GRACE exploits an abstract representation of the social environment and consider the concurrent contribution of behavioural features such as proximity and gaze. However, it is designed with modular components so that additional behavioural features can be considered in the future. This work models each participant in the interaction with an intrinsic Willingness to Engage (WtE) that can be conveyed by their behaviours and perceived by the peer. Instead of modelling WtE symmetrically (as suggested in [205]), we consider the individual contribution to the interaction in an isolated fashion and compute the engagement as a combination of the WtE.

The model introduces the concept of Relevant Feature (RF) as a measurable behaviour of a social agent that can convey social semantics. In [125], authors refer to it as "gestures", but here we want to stress that these cues are relevant for social interaction and can include body parts different from arms and hands. Such features communicate to the partner the degree to which each social agent would like to interact in an intuitive fashion [99]. The presented model builds a metric that exploits the relative behaviour of social agents' body parts. In contrast to other metrics for engagement such as [49] or [111] that model engagement as black boxes, we consider Generalized Recognition of Agent Contribution to Engagement (GRACE) explainable as its output directly involve various values from the identified *Relevant Features*. We do not focus on how the information about the social scene is given to the robot, but rather on how it can be used to understand the interaction from an abstract representation. The model outputs a real-time assessment of the engagement between the participants, linked to the similarity of their WtE.

7.2.1 Methods

Let us define the model that estimates the engagement between two social agents (human and robot). Let n be the number of RF identified for the model, such as proximity, mutual gaze, facial expressions, and others. Let be R the robot, and H the human, we can now express the WtE of a social agent (W_{Agent}) as a vector of size n where each entry is a function $w_{Agent,i}(x) : R \to \{0, k\}$ with $k \in \mathbb{R}^+$ that models the ith feature.

The function $w(x)$[5] belongs to the set A as:

$$A = B \cap \{w(x) = w(-x)\} \tag{7.1}$$

where B is given by:

$$B = \mathscr{S} \cup \{w = \frac{1}{(x^{2n} + 1)} : n \in \mathbb{N}\} \tag{7.2}$$

and $\mathscr{S}$ denotes the Schwartz Space or Schwartz function Space [33]. Notice that we empirically defined the set A, here are some examples of functions

[5] In equations 7.1, 7.2, 7.3 and 7.4, the pedices $_{Agent,i}$ are neglected for sake of readability.

that respect these constraints:

$$w = \frac{1}{((x+\mu)^{2n}+1)} \qquad \forall n \in \mathbb{N} \tag{7.3}$$

$$w = e^{-(x-\mu)^2} \qquad \forall x \in \mathbb{R} \tag{7.4}$$

The scalar μ applies a horizontal shift of the function w. All the functions in the set A can be used as $w(x)$, however, without loss of generality, we select (eq. 7.4) to continue the derivation.

For the human agent, considering the mean $\mu_{H,i}$, we can express $w_{H,i}$ as

$$w_{H,i} = e^{-(x-\mu_{H,i})^2} \tag{7.5}$$

Similarly, $w_{R,i}$ for the robot agent can be expressed as:

$$w_{R,i} = e^{-(x-\mu_{R,i})^2} \tag{7.6}$$

The model returns the partial engagement of the ith feature as the intersection of the agents' WtE, according to the following equation:

$$m_i | w_{R,i} = w_{H,i} \tag{7.7}$$

The scalar m_i is named *partial engagement of the ith feature* ($m_i \in (0,1)$) and provides a measure of how much the agents are engaged on that feature. To this purpose, we firstly need to craft the ideal configuration per each social feature. For example, the ideal configuration for mutual gaze is when social agents are looking at each other, whereas for proximity the ideal configuration can be when the interpersonal distance is at a fixed known value. The mean values $\mu_{H,i}$ and $\mu_{R,i}$ tend to each other when the social agents are fully engaged on the ith feature. These values tend to be different when the agents are fully disengaged on the ith feature.

The properties of the functions of type (eq. 7.3) and (eq. 7.4) guarantee that a solution to eq. 7.7 is unique if $\mu_{H,i} \neq \mu_{R,i}$. This constraint introduces the case for which the curves overlap completely ($\mu_{H,i} = \mu_{R,i}$), in such a case m_i is artificially set to its highest value (1.0).

Let us introduce M as the vector of size n where the ith component is m_i and refers to the ith feature. This vector provides the information, feature-wise, of how much the agents are engaged. Now, we have to re-

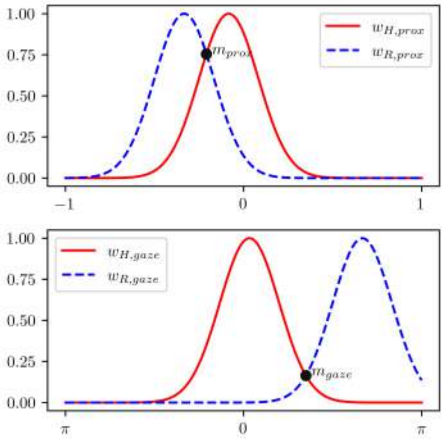


Figure 7.5. Example of Willingness to Engage (WtE) per each feature.

trieve an instantaneous estimation of the engagement. If all the features in the model contribute equally to the engagement, a simple average of this vector provides the value of the instantaneous engagement $m(t)$; otherwise, if features contribute to the interaction non-evenly, a weighted average of M is performed to obtain the value of the engagement. However, social interactions happen dynamically, and a single data point can poorly provide useful information. For this reason, we consider the last $\tau > 0$ instances for assessing the engagement as:

$$\hat{m} = \frac{1}{\tau} \int_{t-\tau}^{t} m(t)dt \tag{7.8}$$

Therefore, $\hat{m}$ is the engagement a robot shall measure when about to initiate an interaction. The validity of this metric can be found in its flexibility and transparency on how much each social feature contributed to its outcome. The metric exploits the semantics of social environments, assuming that there are specific social configuration that facilitate interactions.

The approach described so far is modelling with Gaussian functions various social behaviours. Gaussian functions have already been employed for addressing similar challenges. For instance, in [195] authors model engagement as a multidimensional Gaussian function that varies according to participants' behaviours. It is reasonable to assume that there are some specific behaviours that, when present, can alter interactions. A similar assumption is made in [12] where authors run a time analysis on the recorded interaction data to individuate synchronous events that occurred when agents synchronised their actions.

7.2.2 Relevant Features

The model requires defining the set of Relevant Feature (RF) for the interaction as $\Omega_{RF} = \{RF_0, RF_1, \ldots, RF_n\}$. Each RF_i is characterised by a communication channel (either verbal, non-verbal or para-verbal), a weight that indicates how much it influences the engagement and a mathematical model that maps the semantic of the feature to a scalar value. Proximity and mutual gaze are derived in the following subsections with their respective RF model.

Proximity

We define the values of $\mu_{H,prox}$ and $\mu_{R,prox}$ such that the respective proximity functions ($w_{H,prox}$ and $w_{R,prox}$) overlap when users are at a defined interpersonal social distance. This can also be seen as the optimal configuration for the proximity feature. $|P_H|$ and $|P_R|$ are the magnitudes of the position vectors of the human and robot, respectively. Therefore, their interpersonal distance is defined as $d = |P_R - P_H|$.

$$\mu_{H,prox} = d + \epsilon \tag{7.9}$$

$$\mu_{R,prox} = |P_R| \tag{7.10}$$

Inspired by [96], we consider ϵ as the ideal configuration (i.e., a constant) for which agents are fully engaged when at distance ϵ from each other. These can be substituted in (eq. 7.5) and (eq. 7.6), next the engagement of the proximity feature (m_{prox}) can be obtained from (eq. 7.7). It can be seen that when the social agents are at ϵ distance from each other, m_{prox} is at the maximum, and if the interpersonal distance changes, the mutual engagement of the proximity feature decreases. Authors in [133] show that the preferred interpersonal distance to a robot can differ based on the user (dis)likeability of the robot's behaviour. Moreover, in order to tackle personalised interaction, proximity preferences can also vary and these cases would simply require adjusting ϵ on the fly.

Mutual gaze

Given the strict relation between gaze and head pose, we assume at this state that their reference frame shares a common origin. Considering the gaze of each social agent defined with respect to a common *world* reference frame, where $G_H \in \mathbb{R}^6$ is the human's gaze pose and $G_R \in \mathbb{R}^6$ is the robot's gaze pose. We can denote the first entry of each pose as the gaze directions $\hat{g}_H$ and $\hat{g}_R$, and then the *ideal* mutual gaze between the two agents as $\hat{g}_H + \hat{g}_R = 0$. The gazes have opposite signs since the ideal configuration for mutual gaze happens when agents are in a face-to-face condition. This definition follows closely the one provided in Admoni and Scassellati [2].

$\overrightarrow{RH}$ and $\overrightarrow{HR}$ are respectively the vectors pointing from the robot gaze to the (origin) of the human gaze, and from the human gaze to the (origin) of the robot gaze. Notice that the equivalence $\overrightarrow{RH} = -\overrightarrow{HR}$ holds. If the social agents are not in the *ideal* mutual gaze condition, the angles θ_H and θ_R differ from zero, and they measure how "far" a social agent's gaze is from the ideal gaze (towards the partner)" (see Figure 7.6).

With this model, we can also express the condition in which only one of the social agents is looking at the partner, but is receiving no gaze in return.

The angles θ_R and θ_H hold a semantic meaning in the gaze of each social agent. The final step requires substituting these values in (eq. 7.5) and (eq. 7.6) respectively using:

$$\mu_{H,gaze} = tan(\frac{\theta_H}{4}) \tag{7.11}$$

$$\mu_{R,gaze} = -tan(\frac{\theta_R}{4}) \tag{7.12}$$

The tan functions aim to linearize and scale the angular values of $\theta \in (0, 2\pi)$ to real values $\mu_{Agent,gaze} \in \mathbb{R}$.

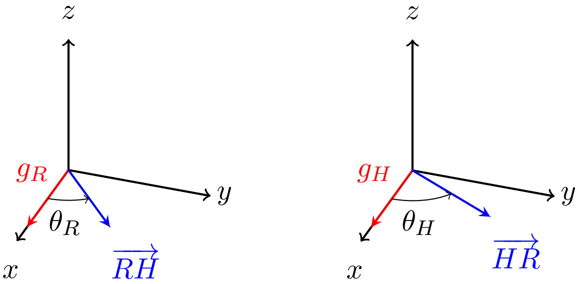


Figure 7.6. Examples of θ_{Agent}. Each angle measures the error between where an agent is gazing and where in space the other participant is located.

This formulation bisects the plane in which $w_{H,gaze}$ and $w_{R,gaze}$ exist, on the origin (different pre-multiplication signs). When users are gazing at each other, the arguments of each function tend to zero by centring both $w_{H,gaze}$ and $w_{R,gaze}$ on the origin. If one of the agents' gaze diverges from the partner, the underlying $w_{Agent,gaze}$ will shift to the proper semi-plane, lowering the value of the intersection (m_{gaze}).

Overall, this approach has computational time cost of $\mathscr{O}(n^2)$ where n is the number of RF. Note that no assumptions are made on the cost for acquiring the RF as this is out of scope in this work.

Validation

Proposing a novel metric for measuring engagement in HRI is challenging and require comparing it with the available metrics in the literature. This metric positions itself as usable within the *initiate* transition of SISM. Therefore, it can only be compared with metrics that follow a similar rationale.

Engagement can be tailored to the interaction context, to the amount of people expecting to interact with the robot, and variables that might limit its generalizability. A critical aspect of validating a novel metric for engagement in HRI is the lack of a unified definition and ground truth data entailing this concept. Despite the limitation being known in the community [182, 141], for the sake of validating GRACE we make few assumptions.

First, the metric presented here positioned itself for measuring the first instances of an interaction. A data source that might fit this scope is found in the UE-HRI dataset [23], as three independent annotators were employed to label it as participants were free to enter or leave the interaction with the robot at will. This characteristic allows modelling the interactions available in the dataset with the SISM.

Second, the metric presented by Del Duchetto *et al.*[49] consists of a regression model (using Convolutional Neural Network (CNN) and Long-Short Term Memory (LSTM) networks) that transforms into a single scalar variable the set of labels given to the annotators in the UE-HRI dataset. This one is handled as it would be ground truth data.

The GRACE model defined thus far need three parameters to be set namely: ϵ, w_{prox}, and w_{gaze}. In our experimental setup, these parameters can be adjusted in real time. With this in mind it is possible to run GRACE alongside the one developed in [49], modify the parameters of GRACE and evaluate how closely the outputs of the metrics are. For doing so, we developed a system of interconnected docker containers to 1) run the grace model, 2) run the model of [49], 3) feed synchronously the same dataset, and 4) aggregate the result for comparing the performances. The advantages of docker containers in this scenario are numerous. With the rationale of allowing peers to replicate these results, we fed both models with a subset of the publicly available dataset UE-HRI [23]. The subset is selected by considering *rosbags* that 1) involve only one person interacting with the robot per time (labelled as *mono* in [23]) and 2) are reliable according to [197] with known standard deviation ($STD = 2.0$).

The subset is summarised in Table 7.1 and is obtained by 1) considering only interactions in which only one person is involved and kept interacting with the robot until the end of the planned interaction (*end-phase* in [23]), and 2) considering the elements of the dataset (*rosbags*) with reliable content according to [197].

Dataset collection

The system is implemented using `Docker Compose` to manage and coordinate distinct services, each encapsulated within its own Docker container. The architecture consists of several services, each interacting via a common ROS ecosystem:

- **roscore**: This container initializes the ROS master, which acts as the central coordination hub, managing communication between all other nodes. The ROS master is assigned a static Internet Protocol (IP) address (`172.21.0.2`) to ensure reliable access by other containers.

- **play_rosbag**: The `play_rosbag` container handles the playback of pre-recorded ROS bag files (see Table 7.1), allowing the system to simulate robotic experiments by replaying sensor data. It depends on the `roscore` container for communication and is mapped to a specific directory (`./play/bags`) to facilitate access to the bag files. The service sets essential environment variables (`ROS_MASTER_URI` and `ROS_HOSTNAME`) to ensure seamless communication with the ROS master.

- **compare**: The `compare` container is responsible for performing the comparison of the metrics. It stores in a `.csv` file the outputs of various metrics along with the parameters of GRACE.

- **run_metric** (*eng_del*, *eng_grace*): Each container is responsible for computing one engagement metric (metric from Del Duchetto *et al.*[49] is computed in *eng_del*) and is configured with its own static IP address to interact with the ROS network. These containers rely on the `play_rosbag` service to process data streamed from the playback of the ROS bag files.

When considering the perspective on how the data flows when collecting the dataset, Figure 7.7 represents with rectangles individual docker containers, and in oval the input set of GRACE parameters. The input parameters θ are randomized during the data collection.

Table 7.1. Subset of rosbags from UE-HRI datasets for performing metrics' benchmark according to their reliability.

user315_2017-02-10	user66_2017-05-12	user1_2017-03-03	user218_2017-02-09
user36_2017-03-13	user8_2017-01-31	user108_2017-03-15	user23_2017-01-20
user4_2017-02-17	user350_2017-04-13	user14_2017-06-14	user230_2017-01-25
user555_2017-04-14	user53_2017-01-31	user16_2017-01-23	user279_2017-04-26
user60_2017-02-20	user68_2017-01-26	user184_2017-02-03	user28_2017-01-31
		user201_2017-02-03	user191_2017-03-16

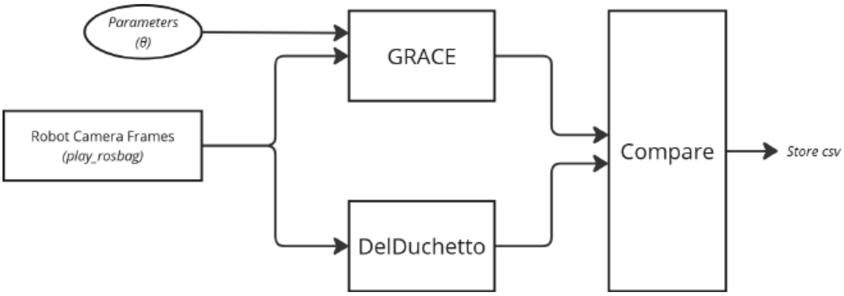


Figure 7.7. Diagram of the docker containers as the dataset is collected.

Figure 7.8 shows the diagram of services' dependencies of docker containers. Lines in solid black define dependent services. For instance, the service "play_rosbags" depends on "roscore". Dashed lines connect services to the custom-made shared network "ros_network". The service "compare" is used to create a dataset in the folder "./csv".

The services in bold run each a different engagement metric. Future works that aim to develop and benchmark an engagement metric with others might use an architecture like this one. For instance, the service called "eng_other" is just a placeholder for other containerised engagement metrics that highlights dependencies and network communication.

The obtained dataset holds $N = 1.197.002$ entries and proceed defining the optimization problem.

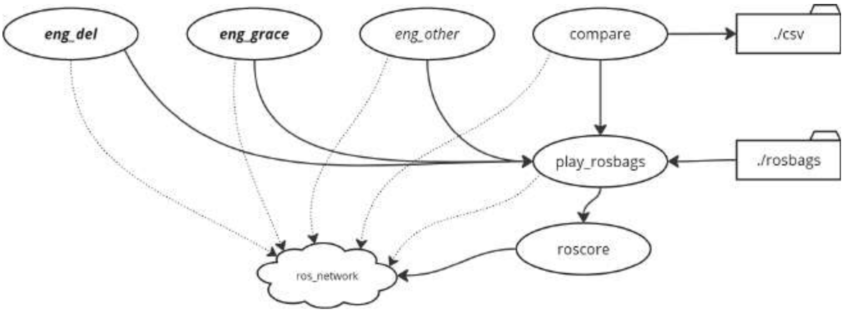


Figure 7.8. Diagram of services' dependencies of docker containers.

Optimization Problem

The optimization problem is formulated to minimize the Mean Squared Error (MSE) between the output of two engagement metrics. The parameters to be optimized are ϵ, w_{prox}, and w_{gaze}. The problem can be described as follows:

Given a dataset $\mathscr{D}$ with columns representing the variables ϵ, w_{prox}, w_{gaze}, and engagement values from two systems, eng_del and $grace\_eng$, the goal is to find optimal values for the parameters $\theta = \epsilon$, w_{prox}, and w_{gaze} that minimize the following objective:

$$\min_{\theta} \sum_{i=1}^{N} (eng\_del_i - eng\_grace_i(\theta))^2$$

where:

- n is the number of data points in the dataset
- eng_del_i represents the output from [49], and $eng\_grace_i(\theta)$ represents the output of GRACE as a function of θ

Through this method, we aimed to optimize the parameters effectively to minimize the difference in engagement values across the systems.

We relaxed the time semantics in this problem and concentrate on finding θ using the library Optuna[6]. Optuna is an automatic hyperparameter optimization software framework that can tackle problems like this one. The optimization process is performed using Optuna's default sampler, and early stopping is implemented if the objective does not improve over a set number of trials. The optimal parameters are the ones that minimize the overall MSE.

For the sake of improving reproducibility of this result, these are the parameters given as input to Optuna.

- $max\_trials = 3000$
- $patience = 1000$
- $min\_delta = 0.00001$

[6] optuna.readthedocs.io/

max_trials defines the maximum amount of trials, *patience* controls how many consecutive trials the optimization will run without improvement before it stops early, and *min_delta* sets the minimum improvement threshold for the objective function between trials.

7.2.3 Results

We split the dataset by considering 80% of it as training set and 20% as test set. We solved the optimization problem on the training set and validate its results using the test set.

The optimization process yielded a set of optimal parameters that minimize the MSE for the engagement metric. The parameters are as follows:

- Proximity parameter ($\epsilon = 0.91$)
- Proximity Weight ($\mathrm{w}_{prox} = 0.21$)
- Gaze Weight ($\mathrm{w}_{gaze} = 0.79$)
- Best MSE$= 0.1883$:

ϵ represents the interpersonal distance (in meters) that allows the two metrics to be as similar as possible. We can consider our result as plausible since an ϵ of about 1 meter falls within the social space range as defined in [71]. w_{prox} and w_{gaze} are the weights attributed to each *RF* when computing the output of GRACE.

Testing these optimal parameters on the unused part of the dataset (the test set) resulted in achieving a MSE of 0.1872.

When examining **RQ3.1** about how to measure engagement in case of non-verbal behaviours, we resorted to the initial phases of interactions. Meaning that, a feasible way of measuring it is by including for the computation of the engagement only the communication channels available. For example, when a person is walking towards a robot in a hall, non-verbal behaviours are available via the onboard sensors even if the person is not yet directly in front of the robot. This reflects situations in which 1) interactions are likely to start and 2) non-verbal behaviours are available to the robot.

The results from the optimization problem show that gaze has a greater impact with respect to proximity when assessing engagement. This outcome is informative when considering the question “To what extent, if any, gaze and proximity affect engagement?” (**RQ3.2**). The study contributes to this question as the model GRACE shall weight proximity for around 22.70% while gaze for 77.30%. The moderate weighting indicates that proximity is an important, but not dominant, factor in determining engagement. These results emphasize the importance of gaze behaviour over proximity in the engagement model of [49], with gaze contributing the majority of the predictive power. A similar result when comparing gaze and proximity was already highlighted by the study presented in Section 5.1.

Finally, the low value of MSE indicates that the selected combination of parameters effectively minimizes the discrepancy between the predicted and actual engagement values, resulting in a robust and accurate model.

Implementation

The model with mutual gaze and proximity as Relevant Features is implemented as a Python3.8 package. The software architecture follows the principle of abstract factory design patterns, so contributing to the project with a new relevant feature can be straightforward. The expected functionality of the software is tested with a set of unit tests that cover up to 90% of the implementation.

The code is designed following the abstract factory design pattern so that other features can be added in the future by simply inheriting from the designed base class. Figure 7.9 shows the class diagram of the implemented software.

A wrapper for ROS noetic is also available and is adapted to be compliant with the interfaces[7] of the project ROS4HRI [130].

To foster the reproducibility of the results and allow peers to benchmark additional metrics on engagement, containerization is published in the form of the popular framework Docker, via text files with instructions for building the binaries given the source code i.e., Dockerfile.

Docker is a platform that enables the creation of lightweight, portable containers that encapsulate software, libraries, and dependencies. This

[7] https://github.com/ros4hri/

technology has proven particularly useful in fields such as robotics, where complex software ecosystems are often required to interact seamlessly [112, 128].

One of the primary advantages of using Docker is its ability to ensure reproducibility across different environments. It is common to utilize various tools, such as ROS nodes, data processing pipelines, and Machine Learning (ML) models.

In this regard, each researcher tackling how to assess engagement in HRI might use different software or hardware for answering their research questions. Once the literature shows various approaches and implementations of tools that can assess engagement, the problem that naturally comes is: Can these metrics be compared? Which metric shall be preferred? The necessity for comparing existing metrics increases, and we present an approach for this challenge. In particular, the implementation of [49] is hosted on github[8] and uses specific versions of dedicated libraries such as Tensorflow 1.14 and Python interpreter (2.∗). Their software loads and uses a pre-trained ML model on the available Graphics Processing Unit (GPU).

Another approach to assessing engagement in HRI is given by Love *et al.*[111] and their implementation differ drastically from the one in [49]. Each engagement metric is isolated in an individual docker container and the ROS communication was established via exploiting the docker network

[8] https://github.com/LCAS/engagement_detector

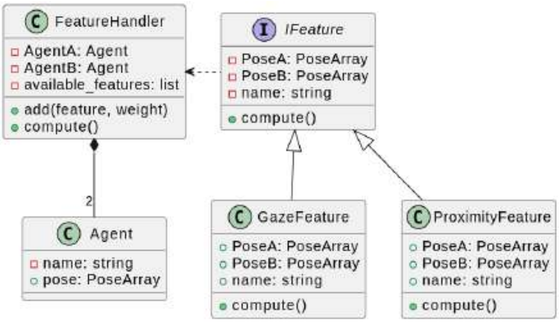


Figure 7.9. Class diagram of the Generalized Recognition of Agent Contribution to Engagement (GRACE) implementation.

functionalities.

The goal is to compare the output of GRACE with the one of another engagement metric, e.g., the one in [49]. Two steps are needed to achieve the goal. First, we collected a dataset with the synchronised assessment of engagement of various metrics while randomly changing the parameters of GRACE. Second, we define an optimization problem on the dataset on a similarity metric between the outputs of GRACE and the one from [49].

Limitations While the implemented model demonstrates promising capabilities in measuring engagement through mutual gaze and proximity, several limitations must be acknowledged: First, the sensitivity of the parameters: ϵ, w_{prox}, and w_{gaze}. Although these parameters can be adjusted in real-time, their optimal values may vary across different contexts or user interactions, potentially affecting the model's generalizability and robustness. The benchmarking of the GRACE model was performed using a subset of the UE-HRI dataset [23]. This subset was limited to specific scenarios (e.g., interactions involving a single person) and may not encompass the full diversity of HRI. As a result, the performance of the model may not accurately reflect its applicability in more complex or varied environments. Another limitation can be found in aiming to replicate the performances of another metric that used expert annotators for their labelling. This solution, despite effective for our use-case, inherently depends on the potential biases of the annotators. Learning the parameters based on different sources will be addressed in future works. The metric currently only considers a single person interacting with a robot. In future works we can extend this to interactions with more people in a group. The model's ability to assess engagement is compared with existing metrics based on a specific rationale (i.e., focusing on the initiation phase of interactions). This comparison limits the scope of evaluation and may overlook other important engagement metrics that could provide additional insights into the quality of interactions. In conclusion, while the GRACE model offers a valuable contribution to measuring engagement in HRI, addressing these limitations will be crucial for enhancing its effectiveness and applicability in diverse real-world contexts.

Overall, this chapter has explored the overarching **RQ3** as “How can engagement be measured in HRI” but tackled how to model and measure engagement (**RQ3.1**) and the impact of gaze and proximity on it (**RQ3.2**). Results indicate the greater impact gaze has with respect to interpersonal distance when evaluating engagement. Improving research in HRI also means allowing researchers to easily compare and benchmark available solutions. This direction is envisioned by the community, and several efforts have been made to facilitate the way we develop software in various robot platforms.

Chapter 8

Conclusions

> *If you only do what you can do, you will never be more than you are now.*
>
> Master Shifu - Kung Fu Panda 3

This thesis investigates how non-verbal robot behaviours can be used to model spontaneous interactions with humans. It is an oxymoron to model *something* that aims to be spontaneous; however, robots must embed clear instructions when finally deployed in our social environments.

In this sense, a spontaneous interaction is considered to happen in unscripted yet intentional ways. These sorts of interactions are expected to happen regularly with social robots employed for autonomous operations within social environments. We refer to this type of interaction as spontaneous Human-Robot Interaction (HRI).

Starting from gaps in the literature that underscore the urge for operationalising *context models* and ways for continuously assessing the robots' surroundings as *context recognition system*, in Chapter 3 we propose the Spontaneous Interaction State Machine (SISM) model, which emphasises the importance of context and interaction state in understanding social behaviours.

The urge for an approach like this is accentuated by the daily interactions social robots are expected to conduct and the dynamic evolution of our social contexts. The appearance and communication capabilities of several robots have been investigated. In particular, the following robots have been used for conducting the reported user studies: ARI from Pal Robotics[1], Tiago also from Pal Robotics, Pepper from SoftBank Robotics[2], *ClassMate* from Protom Robotics[3], Turtlebot2, and Turtlebot4[4].

The studies presented in this thesis orbit around how a robot can behave as to purposefully start and maintain interactions. In other words: "How should robots operate within social environments?". This question introduces the thesis and serves to introduce the three overarching Research Questions (RQs) to study robots' ability to 1) display social cues, 2) purposefully use social cues, and 3) measure interactions. Regarding "How can robots display social cues?" (**RQ1**), the main contributions are as follows:

- a user study investigating how simple social cues from regular vehicles can be transferred to a standard Autonomous Mobile Robot (AMR) available in Section 4.1.
- a user study investigating how complex social cues like emotions can be displayed by a non-humanoid social robot in Section 4.2.

When focusing on "How can robots purposefully use social cues in spontaneous HRI?" (**RQ2**), the contributions are reported in both Chapter 5 and Chapter 6. The contributions that tackle how robots can use social cues to start interactions (**RQ2.1**) are:

- a user study investigating how non-verbal behaviours can be used by a humanoid robot spontaneously approaching a person in a hall, highlighting the role of gaze for signalling the willingness to interact in Section 5.1.
- a user study in which a social robot acting as a bartender can use its social cues to modify the interaction context in Section 5.2.
- a user study investigating how complex social cues like emotions can be used by an AMR approaching a standing human in Section 5.3.

[1] pal-robotics.com
[2] softbankrobotics.com
[3] protomrobotics.com
[4] turtlebot.com

The contributions that tackle how robots can use social cues to influence interactions (**RQ2.2** and **RQ2.3**) are:

- a user study investigating how different robot's communication styles can influence user performance in a game scenario in Section 6.1.

- a user study investigating how robot's emotion-adaptive proxemics behaviours can be used during a spontaneous conversation in Section 6.2.

The studies are interconnected by the rationale presented in Chapter 3 about Spontaneous Interaction State Machine (SISM). This suggests that spontaneous SISM can be modelled with a Finite State Machine (FSM) to establish if the robot is currently involved in a social interaction, if it is about to start one, or if it has just terminated one.

The underlying assumption is that social robots shall be capable of "reading the room" and measure interactions (**RQ3**). On this topic, Chapter 7 proposes lightweight engagement metric called GRACE. This metric is built in an explainable manner and linked to a defined interaction between two features with known social semantics. A set of parameters has to be provided to the metric and influences the output in a non-linear way. These parameters are associated with the semantic of the interaction, they either define the ideal interpersonal distance to interact (ϵ) and the relative weights of the available features (gaze and proximity).

An optimization problem is introduced to find the set of parameters such that the performances of Generalized Recognition of Agent Contribution to Engagement (GRACE) can mirror the ones from another metric already present in the literature [49]. The results indicate that gaze weights significantly more (around 79%) with respect to proximity (around 21%), and that the ideal interpersonal distance (ϵ) is of about $1m$ from the robot.

This finding underscores the role of gaze as a key social cue that can strongly influence users' perception of the robot as an intelligent social actor. Furthermore, this aligns with the recommendations outlined in [210], which support the prioritisation of facial expressions, eye gaze, and purposeful movement in the development of trustworthy robots that foster users' perception of them as intelligent social actors. The importance of gaze as a social cue is also highlighted by the results in Section 5.1, as participants in the user study were able to capture the social intention of a humanoid robot approaching them, mostly thanks to its gazing behaviour.

A tool to improve the reliability of *rosbag* datasets was developed and was used to filter unreliable data for the optimization problem. This step highlights the importance of reliable datasets for improving interaction metrics for research in HRI. This methodological contribution aim to improve the rigour and reproducibility of research in HRI.

With the same goal and to foster the standardisation of research techniques, the implementations used for the user studies and the tools developed are publicly available[5]. Again on the methodological contributions, an approach for comparing engagement metrics that exploit containerisation techniques is presented. This is used for running simultaneously and on the same machine heterogeneous software packages while carefully controlling for desired intra processes (e.g. ROS topics).

8.1 Take-aways

This thesis provides new insights into the nature of spontaneous interactions between humans and robots, emphasizing the importance of non-verbal behaviours, adaptive communication styles, and emotional-adaptive behaviours in the design of socially intelligent robots. With the contributions of this thesis in mind, the following take-aways are identified:

- **Spontaneous Interaction Model:** The introduction of the Spontaneous Interaction State Machine (SISM) highlights the necessity for robots to understand context and interaction states. This model is vital for adapting robot behaviour in dynamic social environments, enabling more fluid and natural interactions.

- **Importance of Non-Verbal Cues:** The findings underscore the significance of non-verbal behaviours, particularly gaze, in initiating and maintaining interactions. Gaze serves as a critical social cue that influences users' perceptions of robots as intelligent social actors.

- **Adaptive Communication:** The research emphasises the need for robots to adopt flexible and adaptable communication styles that can

[5] https://github.com/vignif/AnnexThesis

personalise interactions according to individual user characteristics, including emotional states. This adaptability enhances the quality of the interaction and can be framed as a possible operationalisation of Emotional Intelligence (EI).

8.1.1 Limitations

While this research provides valuable insights, several limitations should be acknowledged: Despite the effort for considering ecological validity, the studies were conducted in controlled environments, which may not fully capture the complexities and variability of real-world interactions. Future work should explore diverse settings to validate the findings. Although emotional cues were explored, the nuances of human emotional responses are complex and may not be fully represented in the robotic interactions tested. Additional investigation is required to explore the complexities of emotional-adaptive behaviours in greater depth. The GRACE metric requires careful selection of parameters, which may vary across different contexts and users. Future work should explore adaptive methods for parameter tuning to enhance the metric's applicability.

In summary, this thesis has provided new insights into the nature of spontaneous interactions between humans and robots, highlighting the importance of non-verbal behaviours, adaptive communication styles and emotional-adaptive behaviours in the design of socially intelligent robots. The research findings have significant implications for the future development of social robots capable of interacting with humans in fluid, natural and socially appropriate ways in a variety of real-world contexts.

Acronyms

AI Artificial Intelligence. 11, *Glossary:* Artificial Intelligence

AMR Autonomous Mobile Robot. xii, 2, 12, 17, 40, 42, 46, 53, 73, 74, 81, 82, 136

API Application Programming Interface. 100

BFI Big-Five Inventory. 92, *Glossary:* Big-Five Inventory

CNN Convolutional Neural Network. 30, 125

CRT Cognitive Reflection Test. 89, 93, 95–97, *Glossary:* Cognitive Reflection Test

DL Deep Learning. *Glossary:* Deep Learning

DoF Degrees of Freedom. 21, 48, 66

EI Emotional Intelligence. 10, 26, 47, 86, 106, 139, *Glossary:* Emotional Intelligence

FER Facial Expression Recognition. 26, *Glossary:* Facial Expression Recognition

FSM Finite State Machine. 33, 35, 137

GAM General Aggression Model. xi, 34, 37

GenAI Generative Artificial Intelligence. *Glossary:* Generative Artificial Intelligence

GPU Graphics Processing Unit. 131

GRACE Generalized Recognition of Agent Contribution to Engagement. xiii, 118, 119, 124–126, 128–132, 137, 139, *Glossary:* Generalized Recognition of Agent Contribution to Engagement

HHI Human-Human Interaction. 22, 26, 47, 55, 56, 78, 86, *Glossary:* Human-Human Interaction

HRI Human-Robot Interaction. i, 4, 5, 8–13, 15, 16, 18–20, 22, 25–30, 33–35, 37, 38, 49, 56, 59, 61, 71, 72, 81–83, 85–87, 98–100, 105–111, 115, 117, 118, 124, 131–133, 135, 136, 138, 146, 175, *Glossary:* Human-Robot Interaction

HRIES Human–Robot Interaction Evaluation Scale. xii, xv, 65, 68, 69, 77, 79, 80, 82, *Glossary:* Human–Robot Interaction Evaluation Scale

IP Internet Protocol. 126

LED Light Emitted Diode. 17, 41, 42, 45, 48

LiDAR Light Detection and Ranging. 78

LLM Large Language Model. 85, 100

LSTM Long-Short Term Memory. 7, 30, 125

ML Machine Learning. 131, *Glossary:* Machine Learning

MSE Mean Squared Error. 128–130

NLP Natural Language Processing. 85, *Glossary:* Natural Language Processing

NLU Natural Language Understanding. *Glossary:* Natural Language Understanding

PCB Printed Circuit Board. 40

PCT Perceptual Control Theory. 10, 85

pmd preferred minimum distance. xii, xv, 73–78, 80

PRAM Persuasive Robots Acceptance Model. xiii, 92, 94, 95, 97, *Glossary:* Persuasive Robots Acceptance Model

PSI Perceived Social Intelligence. xiii, 103–105, *Glossary:* Perceived Social Intelligence

RF Relevant Feature. 118, 119, 122, 124, *Glossary:* Relevant Feature

ROS Robot Operating System. 28, 29, 77, 87, 111–115, 117, 125, 126, 130, 131, 138, 148

ROS1 Robot Operating System 1. 111, 114, *Glossary:* Robot Operating System 1

ROS2 Robot Operating System 2. 111, *Glossary:* Robot Operating System 2

RQ Research Question. 9–13, 76, 82, 99, 136, *Glossary:* Research Question

RSB Robotics Service Bus. 28

SAM Self-Assessment Manikin. 103, 105, *Glossary:* Self-Assessment Manikin

SAR Socially Assistive Robot. 26, *Glossary:* Socially Assistive Robot

SISM Spontaneous Interaction State Machine. i, iii, xi, 12, 33, 35–38, 54, 60–63, 73, 81, 94, 100, 105, 118, 124, 125, 135, 137, 138

STD Standard Deviation. 51, 59, 77, 78, 81, 93–96, 103, 113, 115, 117, 125

SVM Support Vector Machine. 31

TDD Test-Driven Development. 9, *Glossary:* Test-Driven Development

UI User Interface. xii, 66, 70

WtE Willingness to Engage. xiii, 2, 118–121

Glossary

Artificial Intelligence The branch of computer science dedicated to creating systems capable of performing tasks that typically require human intelligence, such as reasoning, learning, problem-solving, and perception. 11, 141

Big-Five Inventory A psychological assessment tool designed to measure five major dimensions of personality traits: openness, conscientiousness, extraversion, agreeableness, and neuroticism, providing insights into individual differences in behaviour. 92, 141

Cognitive Reflection Test A short assessment used to measure an individual's ability to override an initial, intuitive response with a more reflective, correct answer. It commonly includes questions that elicit a fast, intuitive answer, which is often incorrect, encouraging respondents to engage in deeper reasoning to find the correct solution.. 89, 141

Deep Learning A subset of machine learning involving neural networks with multiple layers that enable the modelling of complex patterns and representations in large datasets, particularly effective in tasks such as image and speech recognition. 141

Emotional Intelligence The capacity to recognize, understand, and manage one's own emotions and those of others, playing a crucial role in effective communication, empathy, and interpersonal relationships in both humans and robots. 10, 26, 47, 86, 106, 139, 141

Facial Expression Recognition A technology used to identify human emotions by analyzing facial expressions, often applied in human-robot interaction to enable robots to interpret and respond to the emotional states of users.. 26, 141

Generalized Recognition of Agent Contribution to Engagement An operationalisation of engagement for Human-Robot Interaction that depends non-verbal behaviours of interactants. It is designed to inform about the initial moments of an interaction.. xiii, 119, 131, 137, 142

Generative Artificial Intelligence AI models that leverage algorithms to create original content across various media types, including text, images, and audio, by learning patterns from existing data to produce novel outputs. 142

Human-Human Interaction The study of social and communicative behaviours between two or more individuals, emphasizing the psychological, cultural, and contextual factors that influence interpersonal dynamics. 22, 47, 55, 86, 142

Human-Robot Interaction An interdisciplinary field that explores the interactions and relationships between humans and robots, focusing on understanding how humans communicate, cooperate, and coexist with robotic systems in various contexts. i, 4, 15, 33, 49, 56, 85, 107, 135, 142

Human–Robot Interaction Evaluation Scale A standardized scale developed to evaluate the quality and effectiveness of interactions between humans and robots. It assesses various dimensions of HRI, including perceived safety, comfort, trust, engagement, and social presence. This scale is frequently used in studies to quantify user experiences and the acceptability of robotic systems in different applications.. xii, xv, 65, 69, 79, 80, 142

Machine Learning A subset of artificial intelligence that focuses on the development of algorithms and statistical models that enable computers to learn from and make predictions based on data, enhancing performance without explicit programming. 131, 142

Natural Language Processing A field of artificial intelligence focused on the interactions between computers and human languages, encompassing tasks such as language translation, sentiment analysis, and chatbot development. 85, 142

Natural Language Understanding A subfield of artificial intelligence focused on enabling machines to comprehend and interpret human language, including nuances in meaning, context, and intent, facilitating effective human-computer communication. 142

Perceived Social Intelligence A psychology scale to measure perceptions of robots with a wide range of embodiments and behaviours. The scale measures the robot's ability to understand and respond to social cues, display of empathy and awareness of human emotions, engagement in meaningful communication, and adaptation to social contexts and norms. xiii, 103, 104, 143

Persuasive Robots Acceptance Model A questionnaire designed to measure user acceptance and the perceived persuasiveness of robots in various contexts. If assesses factors such as trust, perceived usefulness, ease of use, and intention to interact, providing insights into user attitudes and acceptance levels towards persuasive robots. It is often used in studies that explore the influence of robot behaviour and appearance on human-robot interaction outcomes.. xiii, 92, 95, 143

Relevant Feature A measurable behaviour of a social agent (a human or a social robot) that can be associated to a defined social semantics.. 118, 122, 143

Research Question A clearly defined query that guides the direction of research studies and experiments, serving as a foundation for hypothesis formulation and data collection strategies. 9, 76, 99, 136, 143

Robot Operating System 1 The first generation of the Robot Operating System, a flexible framework for writing robot software. ROS1 provides tools, libraries, and conventions to simplify the task of creating complex and robust robot behaviour across various robotic platforms. It uses a distributed computing model, with nodes communicating through topics, services, and actions.. 111, 143

Robot Operating System 2 The second generation of the Robot Operating System, designed to overcome the limitations of ROS1 and to enhance real-time performance, scalability, and security. It introduces support for DDS (Data Distribution Service) for improved communication, as well as enhanced multi-robot systems and edge computing capabilities. It is compatible with both Linux and non-Linux platforms, making it more versatile for robotics applications.. 111, 143

Schwartz Space or Schwartz function Space A mathematical function space consisting of all functions whose derivatives decrease faster than any polynomial as they approach infinity. The Schwartz space $\mathscr{S}(\mathbb{R}^n)$ consists of all infinitely differentiable functions $f : \mathbb{R}^n \to \mathbb{C}$ such that

$$\sup_{x\in\mathbb{R}^n} |x^\alpha D^\beta f(x)| < \infty$$

for all multi-indices $\alpha, \beta \in \mathbb{N}_0^n$, where $D^\beta = \frac{\partial^{|\beta|}}{\partial x_1^{\beta_1} \cdots \partial x_n^{\beta_n}}$ is the partial derivative operator [33].. 119

Self-Assessment Manikin A picture-oriented survey to measure dominance, arousal and emotional valence upon defined stimuli. 103, 143

Socially Assistive Robot A robot designed to provide assistance to users, particularly in social and therapeutic contexts, by engaging in interactions that promote emotional well-being, learning, or rehabilitation.. 26, 143

Test-Driven Development A methodology in software development that focuses on an iterative development cycle where the emphasis is placed on writing test cases before the actual feature or function is written. It uses a repetition of short development cycles. This process not only helps improves correctness of the code – but also helps to indirectly evolve the design and architecture of the project at hand.. 9, 143

Bibliography

[1] Gregory D Abowd, Anind K Dey, Peter J Brown, Nigel Davies, Mark Smith, and Pete Steggles. Towards a better understanding of context and context-awareness. In *Handheld and Ubiquitous Computing: First International Symposium, HUC'99 Karlsruhe, Germany, September 27–29, 1999 Proceedings 1*, pages 304–307. Springer, 1999.

[2] Henny Admoni and Brian Scassellati. Social eye gaze in human-robot interaction: a review. *Journal of Human-Robot Interaction*, 6(1):25–63, 2017.

[3] Dustin Aganian, Benedict Stephan, Markus Eisenbach, Corinna Stretz, and Horst-Michael Gross. Attach dataset: Annotated two-handed assembly actions for human action understanding. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11367–11373. IEEE, 2023.

[4] Morana Alač, Javier Movellan, and Fumihide Tanaka. When a robot is social: Spatial arrangements and multimodal semiotic engagement in the practice of social robotics. *Social Studies of Science*, 41(6):893–926, 2011.

[5] Johnie J Allen, Craig A Anderson, and Brad J Bushman. The general aggression model. *Current opinion in psychology*, 19:75–80, 2018.

[6] Amir Aly and Adriana Tapus. A model for synthesizing a combined verbal and nonverbal behavior based on personality traits in human-robot interaction. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 325–332, 2013.

[7] Root Analysis. Root analysis. `https://www.rootsanalysis.com/humanoid-robot-market#:~:text=Humanoid%20Robot%20Market%20Overview,the%20forecast%20period%202024%2D2035.`, 2024. Accessed: 2024-10-10.

[8] Antonio Andriella, Ruben Huertas-Garcia, Santiago Forgas-Coll, Carme Torras, and Guillem Alenyà. “i know how you feel”: The importance of interaction style on users’ acceptance in an entertainment scenario. *Interaction Studies*, 23(1):21–57, 2022.

[9] Antonio Andriella, Henrique Siqueira, Di Fu, Sven Magg, Pablo Barros, Stefan Wermter, Carme Torras, and Guillem Alenyà. Do i have a personality? endowing care robots with context-dependent personality traits. *Internation Journal of Social Robotics*, 2020.

[10] Georgios Angelopoulos, Dimitri Lacroix, Ricarda Wullenkord, Alessandra Rossi, Silvia Rossi, and Friederike Eyssel. Measuring transparency in intelligent robots. *arXiv preprint arXiv:2408.16865*, 2024.

[11] Georgios Angelopoulos, Francesco Vigni, Alessandra Rossi, Giuseppina Russo, Mario Turco, and Silvia Rossi. Familiar acoustic cues for legible service robots. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1187–1192. IEEE, 2022.

[12] Salvatore M Anzalone, Sofiane Boucenna, Serena Ivaldi, and Mohamed Chetouani. Evaluating the engagement with social robots. *International Journal of Social Robotics*, 7:465–478, 2015.

[13] Michael Argyle and Janet Dean. Eye-contact, distance and affiliation. *Sociometry*, pages 289–304, 1965.

[14] Simone Arreghini, Gabriele Abbate, Alessandro Giusti, and Antonio Paolillo. A service robot in the wild: Analysis of users intentions, robot behaviors, and their impact on the interaction. *arXiv preprint arXiv:2410.03287*, 2024.

[15] João Avelino, Leonel Garcia-Marques, Rodrigo Ventura, and Alexandre Bernardino. Break the ice: a survey on socially aware engagement for human–robot first encounters. *International Journal of Social Robotics*, 13(8):1851–1877, 2021.

[16] Alexandra Bacula, Jason Mercer, Jaden Berger, Julie Adams, and Heather Knight. Integrating robot manufacturer perspectives into legible factory robot light communications. *ACM Transactions on Human-Robot Interaction*, 12(1):1–33, 2023.

[17] Massimo Banzi and Michael Shiloh. *Getting started with Arduino: the open source electronics prototyping platform.* Maker Media, Inc., 2014.

[18] Kimberly A Barchard, Leiszle Lapping-Carr, R Shane Westfall, Andrea Fink-Armold, Santosh Balajee Banisetty, and David Feil-Seifer. Measuring the perceived social intelligence of robots. *ACM Transactions on Human-Robot Interaction (THRI)*, 9(4):1–29, 2020.

[19] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1:71–81, 2009.

[20] Andrew Barto, Marco Mirolli, and Gianluca Baldassarre. Novelty or surprise? *Frontiers in psychology*, 4:907, 2013.

[21] Aryel Beck, Lola Cañamero, Luisa Damiano, Giacomo Sommavilla, Fabio Tesser, and Piero Cosi. Children interpretation of emotional body language displayed by a robot. In *International Conference on Social Robotics*, pages 62–70. Springer, 2011.

[22] Tony Belpaeme, Paul Baxter, Robin Read, Rachel Wood, Heriberto Cuayáhuitl, Bernd Kiefer, Stefania Racioppa, Ivana Kruijff-Korbayová, Georgios Athanasopoulos, Valentin Enescu, et al. Multimodal child-robot interaction: Building social bonds. *Journal of Human-Robot Interaction*, 1(2), 2012.

[23] Atef Ben-Youssef, Chloé Clavel, Slim Essid, Miriam Bilac, Marine Chamoux, and Angelica Lim. Ue-hri: a new dataset for the study of user engagement in spontaneous human-robot interactions. In *Proceedings of the 19th ACM international conference on multimodal interaction*, pages 464–472, 2017.

[24] Atef Ben-Youssef, Giovanna Varni, Slim Essid, and Chloé Clavel. On-the-fly detection of user engagement decrease in spontaneous human–robot interaction using recurrent and deep neural networks. *International Journal of Social Robotics*, 11:815–828, 2019.

[25] Aniket Bera, Tanmay Randhavane, and Dinesh Manocha. The emotionally intelligent robot: Improving socially-aware human prediction in crowded environments. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

[26] Baris Bilen, Hasan Kivrak Pinar Uluer, and Hatice Kose. Social robot navigation with adaptive proxemics based on emotions. *arXiv preprint arXiv:2401.17663*, 2024.

[27] Dan Bohus and Eric Horvitz. Models for multiparty engagement in open-world dialog. In *Proceedings of the SIGDIAL 2009 Conference: The 10th*

Annual Meeting of the Special Interest Group on Discourse and Dialogue, SIGDIAL '09, page 225–234, USA, 2009. Association for Computational Linguistics.

[28] Margaret M Bradley and Peter J Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59, 1994.

[29] C. Breazeal, C.D. Kidd, A.L. Thomaz, G. Hoffman, and M. Berlin. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 708–713, 2005.

[30] Joost Broekens and Mohamed Chetouani. Towards transparent robot learning through tdrl-based emotional expressions. *IEEE Transactions on Affective Computing*, 12(2):352–362, 2019.

[31] Lluna María Bru-Luna, Manuel Martí-Vilar, César Merino-Soto, and José L Cervera-Santiago. Emotional intelligence measures: A systematic review. In *Healthcare*, volume 9, page 1696. MDPI, 2021.

[32] Hoang-Long Cao, Paola Cecilia Torrico Moron, Pablo G Esteban, Albert De Beir, Elahe Bagheri, Dirk Lefeber, and Bram Vanderborght. "hmm, did you hear what i just said?": Development of a re-engagement system for socially interactive robots. *Robotics*, 8(4):95, 2019.

[33] W Casselman. Introduction to the schwartz space of t\g. *Canadian Journal of Mathematics*, 41(2):285–320, 1989.

[34] Giovanna Castellano, Berardina De Carolis, Nicola Macchiarulo, and Olimpia Pino. Detecting emotions during cognitive stimulation training with the pepper robot. In *Human-Friendly Robotics 2021: HFR: 14th International Workshop on Human-Friendly Robotics*, pages 61–75. Springer, 2022.

[35] Filippo Cavallo, Francesco Semeraro, Laura Fiorini, Gergely Magyar, Peter Sinčák, and Paolo Dario. Emotion modelling for social robotics applications: a review. *Journal of Bionic Engineering*, 15:185–203, 2018.

[36] Oya Celiktutan, Efstratios Skordos, and Hatice Gunes. Multimodal human-human-robot interactions (mhhri) dataset for studying personality and engagement. *IEEE Transactions on Affective Computing*, 10(4):484–497, 2017.

[37] Elizabeth Cha, Yunkyung Kim, Terrence Fong, Maja J Mataric, et al. A survey of nonverbal signaling methods for non-humanoid robots. *Foundations and Trends® in Robotics*, 6(4):211–323, 2018.

[38] Konstantinos Charalampous, Ioannis Kostavelis, and Antonios Gasteratos. Recent trends in social aware robot navigation: A survey. *Robotics and Autonomous Systems*, 93:85–104, 2017.

[39] Chin S Chen, Chia J Lin, and Chun C Lai. Non-contact service robot development in fast-food restaurants. *IEEE Access*, 10:31466–31479, 2022.

[40] Vijay Chidambaram, Yueh-Hsuan Chiang, and Bilge Mutlu. Designing persuasive robots: how robots might persuade people using vocal and nonverbal cues. In *Proceedings of the 7th ACM/IEEE international conference on Human-Robot Interaction*, pages 293–300, 2012.

[41] Stephanie Hui-Wen Chuah and Joanne Yu. The future of service: The power of emotion in human-robot interaction. *Journal of Retailing and Consumer Services*, 61:102551, 2021.

[42] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, et al. What makes a good conversation? challenges in designing truly conversational agents. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–12, 2019.

[43] Terence Cook, Ashlin RK Roy, and Keith M Welker. Music as an emotion regulation strategy: An examination of genres of music and their roles in emotion regulation. *Psychology of Music*, 47(1):144–154, 2019.

[44] Marco Cristani, Giulia Paggetti, Alessandro Vinciarelli, Loris Bazzani, Gloria Menegaz, and Vittorio Murino. Towards computational proxemics: Inferring social relations from interpersonal distances. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, pages 290–297. IEEE, 2011.

[45] Ilenia Cucciniello, Gianluca L'Arco, Alessandra Rossi, Claudio Autorino, Giuseppe Santoro, and Silvia Rossi. Classmate robot: A robot to support teaching and learning activities in schools. In *2022 31st IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2022.

[46] Osvaldo Da Pos and Paul Green-Armytage. Facial expressions, colours and basic emotions. *Colour: design & creativity*, 1(1):2, 2007.

[47] Sylvain Daronnat, Leif Azzopardi, Martin Halvey, and Mateusz Dubiel. Inferring trust from users' behaviours; agents' predictability positively affects trust, task performance and cognitive load in human-agent real-time collaboration. *Frontiers in Robotics and AI*, 8:642201, 2021.

[48] Jan De Houwer. What are implicit measures and why are we using them. *The handbook of implicit cognition and addiction*, pages 11–28, 2006.

[49] Francesco Del Duchetto, Paul Baxter, and Marc Hanheide. Are you still with me? continuous engagement assessment from a robot's point of view. *Frontiers in Robotics and AI*, 7:116, 2020.

[50] Price James Dillard and Lijiang Shen. On the nature of reactance and its role in persuasive health communication. *Communication Monographs*, 72(2):144–168, 2005.

[51] Mark Dingemanse and Simeon Floyd. Conversation across cultures. In *The Cambridge handbook of linguistic anthropology*, pages 447–480. Cambridge University Press, 2014.

[52] Kevin Doherty and Gavin Doherty. Engagement in hci: conception, theory and measurement. *ACM computing surveys (CSUR)*, 51(5):1–39, 2018.

[53] Stefan Ehrlich, Agnieszka Wykowska, Karinne Ramirez-Amaro, and Gordon Cheng. When to engage in interaction—and how? eeg-based enhancement of robot's ability to sense social signals in hri. In *2014 IEEE-RAS International Conference on Humanoid Robots*, pages 1104–1109. IEEE, 2014.

[54] Paul Ekman. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200, 1992.

[55] Rolando Fernandez, Nathan John, Sean Kirmani, Justin Hart, Jivko Sinapov, and Peter Stone. Passive demonstrations of light-based robot signals for improved human interpretability. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 234–239. IEEE, 2018.

[56] Stephen Fiore, Travis Wiltshire, Emilio Lobato, Florian Jentsch, Wesley Huang, and Benjamin Axelrod. Toward understanding social cues and signals in human–robot interaction: effects of robot gaze and proxemic behavior. *Frontiers in Psychology*, 4, 2013.

[57] Kerstin Fischer, Lars C Jensen, Stefan-Daniel Suvei, and Leon Bodenhagen. Between legibility and contact: The role of gaze in robot approach. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 646–651. IEEE, 2016.

[58] Kerstin Fischer, Malte Jung, Lars Christian Jensen, and Maria Vanessa aus der Wieschen. Emotion expression in hri–when and why. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 29–38. IEEE, 2019.

[59] Forbes. World's first robot hotel fires half of its robots. https://www.forbes.com/sites/samshead/2019/01/16/worlds-first-robot-hotel-fires-half-of-its-robot, 2024. Accessed: 2024-09-08.

[60] Mary Ellen Foster, Andre Gaschler, and Manuel Giuliani. Automatically classifying user engagement for dynamic multi-party human–robot interaction. *International Journal of Social Robotics*, 9(5):659–674, 2017.

[61] Mary Ellen Foster, Andre Gaschler, Manuel Giuliani, Amy Isard, Maria Pateraki, and Ronald PA Petrick. Two people walk into a bar: Dynamic multi-party social interaction with a robot agent. In *Proceedings of the 14th ACM international conference on Multimodal interaction*, pages 3–10, 2012.

[62] Antonio Andriella Francesco Vigni and Silvia Rossi. The impact of robot communication style on user task performance. In *2023 I-RIM Conference*, pages 259–261. I-RIM, 2023.

[63] Shane Frederick. Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4):25–42, 2005.

[64] Siva Ganesh and Vanessa Cave. P-values, p-values everywhere!, 2018.

[65] Aimi S Ghazali, Jaap Ham, Emilia I Barakova, and Panos Markopoulos. Poker face influence: Persuasive robot with minimal social cues triggers less psychological reactance. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 940–946, 2018.

[66] Aimi Shazwani Ghazali, Jaap Ham, Emilia Barakova, and Panos Markopoulos. The influence of social cues in persuasive social robots on psychological reactance and compliance. *Computers in Human Behavior*, 87:58–65, 2018.

[67] Aimi Shazwani Ghazali, Jaap Ham, Emilia Barakova, and Panos Markopoulos. Persuasive robots acceptance model (pram): roles of social responses within the acceptance model of persuasive robots. *International Journal of Social Robotics*, 12(5):1075–1092, 2020.

[68] Jean-Christophe Giger, Nuno Piçarra, Patrícia Alves-Oliveira, Raquel Oliveira, and Patrícia Arriaga. Humanization of robots: Is it really such a good idea? *Human Behavior and Emerging Technologies*, 1(2):111–123, 2019.

[69] Nathan Green and Karen Works. Measuring users' attitudinal and behavioral responses to persuasive communication techniques in human robot interaction. In *Proceedings of the 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 778–782. IEEE, 2022.

[70] Hatice Gunes, Frank Broz, Chris S. Crawford, Astrid Rosenthal-von der Pütten, Megan Strait, and Laurel Riek. Reproducibility in human-robot interaction: Furthering the science of hri. *Current Robotics Reports*, 3(4):281–292, Dec 2022.

[71] Edward T Hall. The hidden dimension. *Garden City*, 1966.

[72] Joanna Hall, Terry Tritton, Angela Rowe, Anthony Pipe, Chris Melhuish, and Ute Leonards. Perception of own and robot engagement in human–robot interactions and their dependence on robotics knowledge. *Robotics and Autonomous Systems*, 62(3):392–399, 2014.

[73] Jaap Ham, René Bokhorst, Raymond Cuijpers, David Van Der Pol, and John-John Cabibihan. Making robots persuasive: the influence of combining persuasive strategies (gazing and gestures) by a storytelling robot on its persuasive power. In *Social Robotics: Third International Conference, ICSR 2011, Amsterdam, The Netherlands, November 24-25, 2011. Proceedings 3*, pages 71–83. Springer, 2011.

[74] Jaap Ham, René Bokhorst, Raymond Cuijpers, David van der Pol, and John-John Cabibihan. Making robots persuasive: The influence of combining persuasive strategies (gazing and gestures) by a storytelling robot on its persuasive power. In Bilge Mutlu, Christoph Bartneck, Jaap Ham, Vanessa Evers, and Takayuki Kanda, editors, *Social Robotics*, pages 71–83, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[75] Justin Hart, Reuth Mirsky, Xuesu Xiao, Stone Tejeda, Bonny Mahajan, Jamin Goo, Kathryn Baldauf, Sydney Owen, and Peter Stone. Using human-inspired signals to disambiguate navigational intentions. In *International Conference on Social Robotics*, pages 320–331. Springer, 2020.

[76] Mojgan Hashemian, Ana Paiva, Samuel Mascarenhas, Pedro Alexandre Santos, and Rui Prada. Social power in human-robot interaction: Towards more persuasive robots. In *AAMAS*, pages 2015–2017, 2019.

[77] John Hawksworth, Richard Berriman, and Saloni Goel. Will robots really steal our jobs? an international analysis of the potential long term impact of automation. 2018.

[78] Brandon Heenan, Saul Greenberg, Setareh Aghel-Manesh, and Ehud Sharlin. Designing social greetings in human robot interaction. In *Proceedings*

of the 2014 conference on Designing interactive systems, pages 855–864, 2014.

[79] Marcel Heerink, Ben Krose, Vanessa Evers, and Bob Wielinga. Measuring acceptance of an assistive social robot: a suggested toolkit. In *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication*, pages 528–533. IEEE, 2009.

[80] Fritz Heider and Marianne Simmel. An experimental study of apparent behavior. *The American journal of psychology*, 57(2):243–259, 1944.

[81] Abdelhadi Hireche, Abdelkader Nasreddine Belkacem, Sadia Jamil, and Chao Chen. Newsgpt: Chatgpt integration for robot-reporter. *arXiv preprint arXiv:2311.06640*, 2023.

[82] Guy Hoffman and Xuan Zhao. A primer for conducting experiments in human–robot interaction. *ACM Transactions on Human-Robot Interaction (THRI)*, 10(1):1–31, 2020.

[83] Alexander Hong, Nolan Lunscher, Tianhao Hu, Yuma Tsuboi, Xinyi Zhang, Silas Franco dos Reis Alves, Goldie Nejat, and Beno Benhabib. A multimodal emotional human–robot interaction architecture for social robots engaged in bidirectional communication. *IEEE Transactions on Cybernetics*, 51(12):5954–5968, 2021.

[84] Shanee Honig and Tal Oron-Gilad. Understanding and resolving failures in human-robot interaction: Literature review and model development. *Frontiers in psychology*, 9:861, 2018.

[85] Aike C. Horstmann, Nikolai Bock, Eva Linhuber, Jessica M. Szczuka, Carolin Straßmann, and Nicole C. Krämer. Do a robot's social skills and its objection discourage interactants from switching the robot off? *PLoS ONE*, 13, 2018.

[86] Fortune Business Insights. Fortune business insights. https://www.fortunebusinessinsights.com/humanoid-robots-market-110188, 2024. Accessed: 2024-10-10.

[87] Serena Ivaldi, Salvatore M Anzalone, Woody Rousseau, Olivier Sigaud, and Mohamed Chetouani. Robot initiative in a team learning task increases the rhythm of interaction but not the perceived engagement. *Frontiers in neurorobotics*, 8:5, 2014.

[88] Masaya Iwasaki, Jian Zhou, Mizuki Ikeda, Yuya Onishi, Tatsuyuki Kawamura, and Hideyuki Nakanishi. Acting as if being aware of visitors' attention strengthens a robotic salesperson's social presence. In *Proceedings of*

the 7th international conference on human-agent interaction, pages 19–27, 2019.

[89] Dinesh Babu Jayagopi, Samira Sheikhi, David Klotz, Johannes Wienke, Jean-Marc Odobez, Sebastian Wrede, Vasil Khalidov, Laurent Nguyen, Britta Wrede, and Daniel Gatica-Perez. The vernissage corpus: A multimodal human-robot-interaction dataset. Technical report, 2012.

[90] Eun-Sook Jee, Yong-Jeon Jeong, Chong Hui Kim, and Hisato Kobayashi. Sound design for emotion and intention expression of socially interactive robots. *Intelligent Service Robotics*, 3(3):199–206, 2010.

[91] Oliver P John, Sanjay Srivastava, et al. The big-five trait taxonomy: History, measurement, and theoretical perspectives. 1999.

[92] Michiel Joosse, Manja Lohse, Jorge Gallego Pérez, and Vanessa Evers. What you do is who you are: The role of task context in perceived social robot personality. In *2013 IEEE International Conference on Robotics and Automation*, pages 2134–2139, 2013.

[93] Adam Kendon. *Conducting interaction: Patterns of behavior in focused encounters*, volume 7. CUP Archive, 1990.

[94] Ege Kesim, Tugce Numanoglu, Oyku Bayramoglu, Bekir Berker Turker, Nusrah Hussain, Metin Sezgin, Yucel Yemez, and Engin Erzin. The ehri database: a multimodal database of engagement in human–robot interactions. *Language Resources and Evaluation*, pages 1–25, 2023.

[95] Yunkyung Kim and Bilge Mutlu. How social distance shapes human–robot interaction. *International Journal of Human-Computer Studies*, 72(12):783–795, 2014.

[96] Yunkyung Kim and Bilge Mutlu. How social distance shapes human–robot interaction. *International Journal of Human-Computer Studies*, 72(12):783–795, 2014.

[97] Mark L Knapp, Judith A Hall, and Terrence G Horgan. *Nonverbal communication in human interaction*, volume 1. Holt, Rinehart and Winston New York, 1978.

[98] Woo-Ri Ko, Minsu Jang, Jaeyeon Lee, and Jaehong Kim. Adaptive behavior generation of social robots based on user behavior recognition. In *International Conference on Social Robotics*, pages 188–197. Springer, 2022.

[99] Ivana Kruijff-Korbayova, Johannes Hackbarth, Caspar Jacob, Bernd Kiefer, Matthias Schmitt, Tanja Schneeberger, Tim Schwartz, Hanns-Peter Horn,

and Karsten Bohlmann. Towards intuitive verbal and non-verbal communication for incidental robot-human encounters in clinic hallways. In *International Conference on Human-Robot Interaction*, volume 2020, 2020.

[100] Michał Kuniecki, Joanna Pilarczyk, and Szymon Wichary. The color red attracts attention in an emotional context. an erp study. *Frontiers in human neuroscience*, 9:212, 2015.

[101] Minae Kwon, Malte F Jung, and Ross A Knepper. Human expectations of social robots. In *2016 11th ACM/IEEE international conference on human-robot interaction (HRI)*, pages 463–464. IEEE, 2016.

[102] Chi-Pang Lam, Chen-Tun Chou, Kuo-Hung Chiang, and Li-Chen Fu. Human-centered robot navigation—towards a harmoniously human–robot coexisting environment. *IEEE Transactions on Robotics*, 27(1):99–112, 2010.

[103] Nicole Lazzeri, Daniele Mazzei, and Danilo De Rossi. Development and testing of a multimodal acquisition platform for human-robot interaction affective studies. *Journal of Human-Robot Interaction*, 3(2):1–24, 2014.

[104] Jeannie S Lee, Muhamed Fauzi Bin Abbas, Chee Kiat Seow, Qi Cao, Kar Peo Yar, Sye Loong Keoh, and Ian McLoughlin. Non-verbal auditory aspects of human-service robot interaction. In *2021 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*, pages 1–5. IEEE, 2021.

[105] Seungcheol Austin Lee and Yuhua Jake Liang. Robotic foot-in-the-door: Using sequential-request persuasive strategies in human-robot interaction. *Computers in Human Behavior*, 90:351–356, 2019.

[106] Séverin Lemaignan, Fernando Garcia, Alexis Jacq, and Pierre Dillenbourg. From real-time attention assessment to "with-me-ness" in human-robot interaction. *ACM/IEEE International Conference on Human-Robot Interaction*, 2016-April:157–164, 2016.

[107] Baisong Liu, Daniel Tetteroo, and Panos Markopoulos. A systematic review of experimental work on persuasive social robots. *International Journal of Social Robotics*, 14(6):1339–1378, 2022.

[108] Tianlin Liu and Arvid Kappas. Predicting engagement breakdown in hri using thin-slices of facial expressions. In *Workshops at the thirty-second AAAI conference on artificial intelligence*, 2018.

[109] Diana Löffler, Nina Schmidt, and Robert Tscharn. Multimodal expression of artificial emotion in social robots using color, motion and sound. In

Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, pages 334–343, 2018.

[110] Matthew Lombard and Kun Xu. Social responses to media technologies in the 21st century: The media are social actors paradigm. *Human-Machine Communication*, 2:29–55, 2021.

[111] Tamlin Love, Antonio Andriella, and Guillem Alenyà. Towards explainable proactive robot interactions for groups of people in unstructured environments. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 697–701, 2024.

[112] Francesco Lumpp, Marco Panato, Franco Fummi, and Nicola Bombieri. A container-based design methodology for robotic applications on kubernetes edge-cloud architectures. In *2021 Forum on specification & Design Languages (FDL)*, pages 01–08. IEEE, 2021.

[113] Keith R MacArthur, Kimberly Stowers, and Peter A Hancock. Human-robot interaction: proximity and speed—slowly back away from the robot! In *Advances in human factors in robots and unmanned systems*, pages 365–374. Springer, 2017.

[114] Gianpaolo Maggi, Elena Dell'Aquila, Ilenia Cucciniello, and Silvia Rossi. "don't get distracted!": The role of social robots' interaction style on users' cognitive performance, acceptance, and non-compliant behavior. *International Journal of Social Robotics*, 13:2057–2069, 2020.

[115] Jon K. Maner and Mary A. Gerend. Motivationally selective risk judgments: Do fear and curiosity boost the boons or the banes? *Organizational Behavior and Human Decision Processes*, 103(2):256–267, 2007.

[116] Umberto Maniscalco, Pietro Storniolo, and Antonio Messina. Bidirectional multi-modal signs of checking human-robot engagement and interaction. *International Journal of Social Robotics*, pages 1–15, 2022.

[117] Samuel Marcos-Pablos and Francisco José García-Peñalvo. Emotional intelligence in robotics: A scoping review. In *New Trends in Disruptive Technologies, Tech Ethics and Artificial Intelligence: The DITTET Collection 1*, pages 66–75. Springer, 2022.

[118] Claudia Marinetti, Penny Moore, Pablo Lucas, and Brian Parkinson. Emotions in social interactions: Unfolding emotional experience. *Emotion-oriented systems: The humaine handbook*, pages 31–46, 2011.

[119] Alva Markelius. An empirical design justice approach to identifying ethical considerations in the intersection of large language models and social robotics. *arXiv preprint arXiv:2406.06400*, 2024.

[120] Richard S Marken and Warren Mansell. Perceptual control as a unifying concept in psychology. *Review of General Psychology*, 17(2):190–195, 2013.

[121] Mina Marmpena, Angelica Lim, Torbjørn S Dahl, and Nikolas Hemion. Generating robotic emotional body language with variational autoencoders. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 545–551. IEEE, 2019.

[122] Gerald Matthews and Moshe Zeidner. Emotional intelligence, adaptation to stressful encounters, and health outcomes. 2000.

[123] Derek McColl, Zhe Zhang, and Goldie Nejat. Human body pose interpretation and classification for social human-robot interaction. *International Journal of Social Robotics*, 3(3):313–332, 2011.

[124] Mary L McHugh. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3):276–282, 2012.

[125] David McNeill. Hand and mind1. *Advances in Visual Semiotics*, page 351, 1992.

[126] Ross Mead and Maja J Matarić. Proxemics and performance: Subjective human evaluations of autonomous sociable robot distance and social signal understanding. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5984–5991. IEEE, 2015.

[127] Hanneke KM Meeren, Corné CRJ Van Heijnsbergen, and Beatrice De Gelder. Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences*, 102(45):16518–16523, 2005.

[128] Pedro Melo, Rafael Arrais, and Germano Veiga. Development and deployment of complex robotic applications using containerized infrastructures. In *2021 IEEE 19th International Conference on Industrial Informatics (INDIN)*, pages 1–8. IEEE, 2021.

[129] Paulo Menezes, Joao Quintas, and Jorge Dias. The role of context information in human-robot interaction. In *23rd IEEE International Symposium on Robot and Human Interactive Communication Workshop on Interactive Robots for Aging and/or Impaired People*, page 20, 2014.

[130] Youssef Mohamed and Séverin Lemaignan. Ros for human-robot interaction. In *2021 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 3020–3027. IEEE, 2021.

[131] Lilia Moshkina. Improving request compliance through robot affect. *Proceedings of the AAAI Conference on Artificial Intelligence*, 26:2031–2037, 2012.

[132] Lilia Moshkina, Susan Trickett, and J Gregory Trafton. Social engagement in public places: a tale of one robot. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 382–389, 2014.

[133] Jonathan Mumm and Bilge Mutlu. Human-robot proxemics: Physical and psychological distancing in human-robot interaction. In *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 331–338, 2011.

[134] Venkatraman Narayanan, Bala Murali Manoghar, Vishnu Sashank Dorbala, Dinesh Manocha, and Aniket Bera. Proxemo: Gait-based emotion learning and multi-view proxemic fusion for socially-aware robot navigation. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8200–8207. IEEE, 2020.

[135] Jauwairia Nasir, Barbara Bruno, Mohamed Chetouani, and Pierre Dillenbourg. What if social robots look for productive engagement? *International Journal of Social Robotics*, 14(1):55–71, 2022.

[136] Margot ME Neggers, Raymond H Cuijpers, and Peter AM Ruijten. Comfortable passing distances for robots. In *Social Robotics: 10th International Conference, ICSR 2018, Qingdao, China, November 28-30, 2018, Proceedings 10*, pages 431–440. Springer, 2018.

[137] Margot ME Neggers, Raymond H Cuijpers, Peter AM Ruijten, and Wijnand A IJsselsteijn. Determining shape and size of personal space of a human when passed by a robot. *International Journal of Social Robotics*, 14(2):561–572, 2022.

[138] Ha Quang Thinh Ngo, Thanh Phuong Nguyen, and Hung Nguyen. Investigation on barbot to serve human in public space. In *2018 4th International Conference on Green Technology and Sustainable Development (GTSD)*, pages 300–305. IEEE, 2018.

[139] Jekaterina Novikova and Leon Watts. Towards artificial emotions to assist social coordination in hri. *International Journal of Social Robotics*, 7:77–88, 2015.

[140] Heather L O'Brien and Elaine G Toms. What is user engagement? a conceptual framework for defining user engagement with technology. *Journal of the American society for Information Science and Technology*, 59(6):938–955, 2008.

[141] Catharine Oertel, Ginevra Castellano, Mohamed Chetouani, Jauwairia Nasir, Mohammad Obaid, Catherine Pelachaud, and Christopher Peters. Engagement in Human-Agent Interaction: An Overview. *Frontiers in Robotics and AI*, 7:92, 2020.

[142] Maike Paetzel-Prüsmann, Giulia Perugia, and Ginevra Castellano. The influence of robot personality on the development of uncanny feelings. *Computers in Human Behavior*, 120:106756, 2021.

[143] Panagiotis Papadakis, Patrick Rives, and Anne Spalanzani. Adaptive spacing in human-robot interactions. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2627–2632. IEEE, 2014.

[144] Panagiotis Papadakis, Anne Spalanzani, and Christian Laugier. Social mapping of human-populated environments by implicit function learning. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1701–1706. IEEE, 2013.

[145] Christine E Parsons, Katherine S Young, Else-Marie Jegindoe Elmholdt, Alan Stein, and Morten L Kringelbach. Interpreting infant emotional expressions: Parenthood has differential effects on men and women. *Quarterly journal of experimental psychology*, 70(3):554–564, 2017.

[146] André Pereira, Catharine Oertel, Leonor Fermoselle, Joseph Mendelson, and Joakim Gustafson. Effects of different interaction contexts when evaluating gaze models in hri. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pages 131–139, 2020.

[147] Björn Petrak, Julia G Stapels, Katharina Weitz, Friederike Eyssel, and Elisabeth André. To move or not to move? social acceptability of robot proxemics behavior depending on user emotion. In *2021 30th IEEE international conference on robot & human interactive communication (RO-MAN)*, pages 975–982. IEEE, 2021.

[148] Robert Plutchik. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist*, 89(4):344–350, 2001.

[149] Isabella Poggi. *Mind, hands, face and body: a goal and belief view of multimodal communication.* Weidler, 2007.

[150] Luca Raggioli, Fabio Aurelio D'Asaro, and Silvia Rossi. Deep reinforcement learning for robotic approaching behavior influenced by user activity and disengagement. *International Journal of Social Robotics*, 2023.

[151] Daniel J Rea, Sebastian Schneider, and Takayuki Kanda. " is this all you can do? harder!" the effects of (im) polite robot encouragement on exercise effort. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, pages 225–233, 2021.

[152] James Rehg, Gregory Abowd, Agata Rozga, Mario Romero, Mark Clements, Stan Sclaroff, Irfan Essa, O Ousley, Yin Li, Chanho Kim, et al. Decoding children's social behavior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3414–3421, 2013.

[153] Goldman Sachs Research. Goldman sachs. https://www.goldmansachs.com/insights/articles/the-global-market-for-robots-could-reach-38-billion-by-2035/, 2024. Accessed: 2024-10-10.

[154] Mauricio E Reyes, Ivan V Meza, and Luis A Pineda. Robotics facial expression of anger in collaborative human–robot interaction. *International Journal of Advanced Robotic Systems*, 16(1):1729881418817972, 2019.

[155] Charles Rich, Brett Ponsleur, Aaron Holroyd, and Candace L. Sidner. Recognizing engagement in human-robot interaction. In *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction*, HRI '10, page 375–382. IEEE Press, 2010.

[156] Bear Robotics. Bear robotics. https://www.bearrobotics.ai/, 2024. Accessed: 2024-09-08.

[157] Pudu Robotics. Pudu robotics. https://www.pudurobotics.com/, 2024. Accessed: 2024-09-08.

[158] Alessandra Rossi, Fernando Garcia, Arturo Cruz Maya, Kerstin Dautenhahn, Kheng Lee Koay, Michael L. Walters, and Amit K. Pandey. Investigating the effects of social interactive behaviours of a robot on people's trust during a navigation task. In *Towards Autonomous Robotic Systems (TAROS 2019), Lecture Notes in Computer Science*, pages 349–361, Cham, 2019. Springer International Publishing.

[159] Alessandra Rossi, Marcus M Scheunemann, Gianluca L'Arco, and Silvia Rossi. Evaluation of a humanoid robot's emotional gestures for transparent interaction. In *International Conference on Social Robotics*, pages 397–407. Springer, 2021.

[160] Alessandra Rossi, Mariacarla Staffa, Antonio Origlia, Maria Di Maro, and Silvia Rossi. Brillo: A robotic architecture for personalised long-lasting interactions in a bartending domain. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, pages 426–429, 2021.

[161] Silvia Rossi, Elena Dell'Aquila, and Benedetta Bucci. Evaluating the emotional valence of affective sounds for child-robot interaction. In *Social Robotics*, pages 505–514, Cham, 2019. Springer International Publishing.

[162] Silvia Rossi, Giovanni Ercolano, Luca Raggioli, Emanuele Savino, and Martina Ruocco. The disappearing robot: An analysis of disengagement and distraction during non-interactive tasks. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 522–527, 2018.

[163] Silvia Rossi and Martina Ruocco. Better alone than in bad company: Effects of incoherent non-verbal emotional cues for a humanoid robot. *Interaction Studies*, 20(3):487–508, 2019.

[164] Silvia Rossi, Mariacarla Staffa, Luigi Bove, Roberto Capasso, and Giovanni Ercolano. User's personality and activity influence on hri comfortable distances. In *Social Robotics*, pages 167–177, Cham, 2017. Springer International Publishing.

[165] James A Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161, 1980.

[166] Hanan Salam and Mohamed Chetouani. Engagement detection based on mutli-party cues for human robot interaction. In *2015 international conference on affective computing and intelligent interaction (ACII)*, pages 341–347. IEEE, 2015.

[167] Hanan Salam and Mohamed Chetouani. A multi-level context-based modeling of engagement in human-robot interaction. In *2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG)*, volume 3, pages 1–6. IEEE, 2015.

[168] David A Salter, Amir Tamrakar, Behjat Siddiquie, Mohamed R Amer, Ajay Divakaran, Brian Lande, and Darius Mehri. The tower game dataset: A multimodal dataset for analyzing social interaction predicates. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 656–662. IEEE, 2015.

[169] SM Bhagya P Samarakoon, MA Viraj J Muthugala, AG Buddhika P Jayasekara, and Mohan Rajesh Elara. Adapting approaching proxemics of a service robot based on physical user behavior and user feedback. *User Modeling and User-Adapted Interaction*, 33(2):195–220, 2023.

[170] Jyotirmay Sanghvi, Ginevra Castellano, Iolanda Leite, André Pereira, Peter W McOwan, and Ana Paiva. Automatic analysis of affective postures and body motion to detect engagement with a game companion. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 305–312, 2011.

[171] Satoru Satake, Takayuki Kanda, Dylan F Glas, Michita Imai, Hiroshi Ishiguro, and Norihiro Hagita. How to approach humans? strategies for social robots to initiate interaction. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 109–116, 2009.

[172] Shane Saunderson and Goldie Nejat. Investigating strategies for robot persuasion in social human–robot interaction. *IEEE Transactions on Cybernetics*, 52(1):641–653, 2020.

[173] Michael Schmitz. Concepts for life-like interactive objects. In *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction*, pages 157–164, 2010.

[174] Sarah Schömbs, Jacobe Klein, and Eileen Roesler. Feeling with a robot—the role of anthropomorphism by design and the tendency to anthropomorphize in human-robot interaction. *Frontiers in Robotics and AI*, 10:1149601, 2023.

[175] Jamieson Schulte, Charles Rosenberg, and Sebastian Thrun. Spontaneous, short-term interaction with mobile robots. In *Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No. 99CH36288C)*, volume 1, pages 658–663. IEEE, 1999.

[176] Alessandra Sciutti, Martina Mara, Vincenzo Tagliasco, and Giulio Sandini. Humanizing human-robot interaction: On the importance of mutual understanding. *IEEE Technology and Society Magazine*, 37(1):22–29, 2018.

[177] Konika Sharma, Patrick Jermann, and Pierre Dillenbourg. "with-me-ness": A gaze-measure for students' attention in moocs. *Proceedings of International Conference of the Learning Sciences, ICLS*, 2:1017–1021, 01 2014.

[178] Chao Shi, Michihiro Shimada, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. Spatial formation model for initiating conversation. In *Robotics: science and systems*, volume 11, 2011.

[179] Chao Shi, Masahiro Shiomi, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. Measuring communication participation to initiate conversation in human–robot interaction. *International Journal of Social Robotics*, 7:889–910, 2015.

[180] Moondeep C. Shrestha, Ayano Kobayashi, Tomoya Onishi, Erika Uno, Hayato Yanagawa, Yuta Yokoyama, Mitsuhiro Kamezaki, Alexander Schmitz, and Shigeki Sugano. Intent communication in navigation through the use of light and screen indicators. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 523–524, 2016.

[181] Candace L Sidner, Christopher Lee, Cory D Kidd, Neal Lesh, and Charles Rich. Explorations in engagement for humans and robots. *Artificial Intelligence*, 166(1-2):140–164, 2005.

[182] Alessandra Sorrentino, Laura Fiorini, and Filippo Cavallo. From the definition to the automatic assessment of engagement in human–robot interaction: A systematic review. *International Journal of Social Robotics*, pages 1–23, 2024.

[183] Nicolas Spatola, Barbara Kühnlenz, and Gordon Cheng. Perception and evaluation in human–robot interaction: The human–robot interaction evaluation scale (hries)—a multicomponent approach of anthropomorphism. *International Journal of Social Robotics*, 13(7):1517–1539, 2021.

[184] Matteo Spezialetti, Giuseppe Placidi, and Silvia Rossi. Emotion recognition for human-robot interaction: Recent advances and future perspectives. *Frontiers in Robotics and AI*, page 145, 2020.

[185] Mariacarla Staffa, Massimo De Gregorio, Maurizio Giordano, and Silvia Rossi. Can you follow that guy? In *22th European Symposium on Artificial Neural Networks, ESANN 2014, Bruges, Belgium, April 23-25, 2014*, 2014.

[186] Christopher John Stanton and Catherine J Stevens. Don't stare at me: the impact of a humanoid robot's gaze upon trust during a cooperative human–robot visual task. *International Journal of Social Robotics*, 9:745–753, 2017.

[187] Rebecca Stower, Natalia Calvo-Barajas, Ginevra Castellano, and Arvid Kappas. A meta-analysis on children's trust in social robots. *International Journal of Social Robotics*, 13(8):1979–2001, 2021.

[188] Megan Strait, Lara Vujovic, Victoria Floerke, Matthias Scheutz, and Heather Urry. Too much humanness for human-robot interaction: exposure to highly humanlike robots elicits aversive responding in observers. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 3593–3602, 2015.

[189] Leila Takayama and Caroline Pantofaru. Influences on proxemic behaviors in human-robot interaction. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5495–5502. IEEE, 2009.

[190] Yunus Terzioğlu, Bilge Mutlu, and Erol Şahin. Designing social cues for collaborative robots: the role of gaze and breathing in human-robot collaboration. In *Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction*, pages 343–357, 2020.

[191] Myrthe Tielman, Mark Neerincx, John-Jules Meyer, and Rosemarijn Looije. Adaptive emotional expression in robot-child interaction. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 407–414, 2014.

[192] Elena Torta, Raymond H Cuijpers, James F Juola, and David van der Pol. Design of robust robotic proxemic behaviour. In *Social Robotics: Third International Conference, ICSR 2011, Amsterdam, The Netherlands, November 24-25, 2011. Proceedings 3*, pages 21–30. Springer, 2011.

[193] Rudolph Triebel, Kai Arras, Rachid Alami, Lucas Beyer, Stefan Breuers, Raja Chatila, Mohamed Chetouani, Daniel Cremers, Vanessa Evers, Michelangelo Fiore, et al. Spencer: A socially aware service robot for passenger guidance and help in busy airports. In *Field and Service Robotics: Results of the 10th International Conference*, pages 607–622. Springer, 2016.

[194] Companies using ROS. Companies using ros, 2023. Last accessed: 2023-12-07.

[195] George Velentzas, Theodore Tsitsimis, Iñaki Rañó, Costas Tzafestas, and Mehdi Khamassi. Adaptive reinforcement learning with active state-specific exploration for engagement maximization during simulated child-robot interaction. *Paladyn, Journal of Behavioral Robotics*, 9(1):235–253, 2018.

[196] Francesco Vigni, Antonio Andriella, and Silvia Rossi. Sweet robot o'mine-how a cheerful robot boosts users' performance in a game scenario. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1368–1374. IEEE, 2023.

[197] Francesco Vigni, Antonio Andriella, and Silvia Rossi. A rosbag tool to improve dataset reliability. In *Companion of the 2024 ACM/IEEE international conference on human-robot interaction*, pages 1085–1089, 2024.

[198] Francesco Vigni, Espen Knoop, Domenico Prattichizzo, and Monica Malvezzi. The role of closed-loop hand control in handshaking interactions. *IEEE Robotics and Automation Letters*, 4(2):878–885, 2019.

[199] Francesco Vigni, Dimitri Maglietta, and Silvia Rossi. Too close to you? a study on emotion-adapted proxemics behaviours. In *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*, pages 182–188, 2024.

[200] Francesco Vigni, Alessandra Rossi, Linda Miccio, and Silvia Rossi. On the emotional transparency of a non-humanoid social robot. In *International Conference on Social Robotics*, pages 290–299. Springer, 2022.

[201] Francesco Vigni and Silvia Rossi. Exploring non-verbal strategies for initiating an hri. In *International Conference on Social Robotics*, pages 280–289. Springer, 2022.

[202] Michael L Walters, Kerstin Dautenhahn, Sarah N Woods, Kheng Lee Koay, R Te Boekhorst, and David Lee. Exploratory studies on social spaces between humans and a mechanical-looking robot. *Connection Science*, 18(4):429–439, 2006.

[203] Atsushi Watanabe, Tetsushi Ikeda, Yoichi Morales, Kazuhiko Shinozawa, Takahiro Miyashita, and Norihiro Hagita. Communicating robotic navigational intentions. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5763–5769, 2015.

[204] Adam Waytz, Joy Heafner, and Nicholas Epley. The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of experimental social psychology*, 52:113–117, 2014.

[205] Nicola Webb, Manuel Giuliani, and Séverin Lemaignan. Measuring visual social engagement from proxemics and gaze. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 757–762. IEEE, 2022.

[206] Yalun Wen, Xingwei Wu, Katsu Yamane, and Soshi Iba. Socially-aware mobile robot trajectories for face-to-face interactions. In *International Conference on Social Robotics*, pages 3–13. Springer, 2022.

[207] Johannes Wienke, David Klotz, and Sebastian Wrede. A framework for the acquisition of multimodal human-robot interaction data sets with a whole-system perspective. In *LREC 2012 Workshop on Multimodal Corpora for Machine Learning*. Citeseer, 2012.

[208] Johannes Wienke and Sebastian Wrede. A middleware for collaborative research in experimental robotics. In *2011 IEEE/SICE International Symposium on System Integration (SII)*, pages 1183–1190. IEEE, 2011.

[209] Robots with ROS. Robots with ros, 2023. Last accessed: 2023-12-07.

[210] Kun Xu, Mo Chen, and Leping You. The hitchhiker's guide to a credible and socially present robot: Two meta-analyses of the power of social cues in human–robot interaction. *International Journal of Social Robotics*, 15(2):269–295, 2023.

[211] Mei Ying and Liu Zhentao. An emotion-driven attention model for service robot. In *2016 12th World Congress on Intelligent Control and Automation (WCICA)*, pages 1526–1531. IEEE, 2016.

[212] Atef Ben Youssef, Giovanna Varni, Slim Essid, and Chloé Clavel. On-the-fly detection of user engagement decrease in spontaneous human-robot interaction, international journal of social robotics, 2019. *CoRR*, abs/2004.09596, 2020.

[213] Rui Zhang, Wanyue Jiang, Zhonghao Zhang, Yuhan Zheng, and Shuzhi Sam Ge. Indoor mobile robot socially concomitant navigation system. In *International Conference on Social Robotics*, pages 485–495. Springer, 2022.

[214] Yanxia Zhang, Jonas Beskow, and Hedvig Kjellström. Look but don't stare: Mutual gaze interaction in social robots. In *International Conference on Social Robotics*, pages 556–566. Springer, 2017.

[215] Xuan Zhao and Bertram F Malle. Spontaneous perspective taking toward robots: The unique impact of humanlike appearance. *Cognition*, 224:105076, 2022.

Author's publications

1. Georgios Angelopoulos*, Francesco Vigni*, Alessandra Rossi, Giuseppina Russo, Mario Turco, and Silvia Rossi. Familiar acoustic cues for legible service robots. In 2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), pages 1187–1192. IEEE, 2022.

2. Francesco Vigni and Silvia Rossi. Exploring non-verbal strategies for initiating an hri. In International Conference on Social Robotics, pages 280–289. Springer, 2022.

3. Francesco Vigni, Alessandra Rossi, Linda Miccio, and Silvia Rossi. On the emotional transparency of a non-humanoid social robot. In International Conference on Social Robotics, pages 290–299. Springer, 2022.

4. Francesco Vigni, Antonio Andriella, and Silvia Rossi. Sweet Robot O'Mine - how a cheerful robot boosts users' performance in a game scenario. In 2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), pages 1368–1374. IEEE, 2023.

5. Francesco Vigni, Antonio Andriella, and Silvia Rossi. The impact of robot communication style on user task performance. In 2023 I-RIM Conference, pages 259-261. I-RIM, 2023.

6. Francesco Vigni, Antonio Andriella, and Silvia Rossi. A rosbag tool to improve dataset reliability. In Companion of the 2024 ACM/IEEE international conference on human-robot interaction, pages 1085–1089, 2024.

7. Vasilis Mizaridis*, Francesco Vigni*, Stratos Arampatzis, and Silvia Rossi. Are emotions important? a study on social distances for path planning based on emotions. In 2024 33rd IEEE International Conference on Robot

* co-first authorship

and Human Interactive Communication (RO-MAN). pages 176-181. IEEE, 2024.

8. Francesco Vigni, Dimitri Maglietta, and Silvia Rossi. Too close to you? a study on emotion-adapted proxemics behaviours. In 2024 33rd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). pages 182-188. IEEE, 2024.

9. Francesco Vigni, Esteve Valls Mascaro, Dongheui Lee and Silvia Rossi. Between Task and Social Engagement of a Social Service Robot. *Submitted to* IEEE Robotics and Automation Letters, 2024

10. Francesco Vigni and Silvia Rossi. Measuring the Unmeasurable: Engagement in HRI. *Submitted to* IEEE Robotics and Automation Letters, 2024

Acknowledgements

"You are a citizen of the world" is the mantra that I grew up with (thanks mom). The concept of belonging has followed me throughout the years and places. This PhD journey officially lasted three years; unofficially, it is the result of all the places, people, and robots I interacted with. Each of them in some way motivated me to achieve this.

First, I would like to thank my supervisor **Silvia Rossi** for providing enthusiastic support and great advice. From the beginning, you stood as an outstanding model, providing me with valuable advice and experience that helped foster my development as an independent researcher. Thank you for granting me the autonomy to explore my ideas, for providing the tools to achieve these results, and for guiding me with valuable research directions. Second, I would like to thank **Antonio Andriella**. Thank you for your outstanding feedback and unbounded support that significantly contributed to enduring my motivation throughout these years. You quickly evolved into an important friend and a valuable mentor, constantly supporting me despite the lack of any formality.

Moreover, I would like to thank **Alessandra Rossi** for the valuable contributions and feedback throughout these years, and **Claudia Di Napoli** for the valuable insights and support, especially during the write-up of this thesis. I feel extremely fortunate to have had the opportunity to conduct this research within the European Union's Horizon 2020 research and innovation program called PERSEO. So I deeply thank the fellow students and supervisors part of the project as we, PERSEO, harmonised our efforts to improve the world of HRI and drive the progress of our evolving societies.

I would like to thank **Dongheui Lee** from the Technical University of Vienna in Vienna, Austria, for the opportunity to join her team as a visiting researcher during winter 2023. Thank you for the support, the kindness, and the invaluable contribution to my research.

A special mention goes to my colleagues and friends **Georgios Angelopoulos**, **Pradip Pramanick**, **Dimitri Lacroix**, and **Unai Zabala** with whom I have had the pleasure of sharing numerous experiences both within and beyond the research environment. I would like to thank **Michael Suppa** and the team at Roboception GmbH for significantly sparking my passion for research and highlighting the importance of a healthy work-life balance.

As the leaves of a tree are nurtured by its roots, I am grateful to my family. Your immeasurable support allowed me to flourish and achieve this goal. Thank you to my brother **Alessandro** for the continuous support of all these years and to my mother **Luz Margarita** for fostering in me the desire to continuously improve myself.

Alongside the research, the project gave me the opportunity to discover the poetry that permeates the city of Naples, Italy. Again, my concept of belonging stretches itself to embrace yet another place I can refer to as home. Carving in my heart a sentence that goes beyond the football field: Forza Napoli!
Thank you. Grazie.

Francesco Vigni
Napoli, Italia
31 Ottobre 2024